Statistics 133 Final Project

Christian Alarcio, Ellen Chan, Anais Sidhu, Ruomeng (Michelle) Yang

December 7, 2015

Introduction

Our findings and analysis from this project were led by our hypothesis:

Technology industries show higher growth and higher beta, and therefore lower P/E ratios.

We answered the following questions:

- *What is the relationship between beta and the payout percentage?*
- *What is the relationship between beta and expected growth rate?*

Before we introduce our dataset and preprocessing steps, here is an explanation of some of the financial terminology.

Dataset & Preprocessing

To begin the project, two datasets were downloaded as CSV files from the NYU Stern Business School's data archives. There was an online link that automatically downloaded the data we needed, so no there was no R needed. We start by setting the working directory, then reading the two CSV files. The files were then merged and assigned to the data frame "raw_data".

```
# Set working directory
setwd("~/Documents/UC Berkeley 2015-2016/Statistics 133/projects/final/")

# Extract raw data
beta <- read.csv("raw_data/total_beta.csv", header = TRUE, stringsAsFactors = FALSE)
pe <- read.csv("raw_data/pe_data.csv", header = TRUE, stringsAsFactors = FALSE)

# Combine data from the two files
raw_data <- merge(beta, pe, by = intersect(names(beta), names(pe)))
```

We then inspected the contents of the data frame.

```
# Inspect merged data
head(raw_data)
```

```
##             Industry Number.of.firms Average.Unlevered.Beta
## 1        Advertising              52                   0.83
## 2 Aerospace/Defense               93                   1.06
## 3     Air Transport               22                   0.61
## 4           Apparel               64                   0.86
## 5      Auto & Truck               22                   0.59
## 6        Auto Parts               75                   1.14
##   Average.Levered.Beta Average.correlation Total.Unlevered.Beta
```

```
## 1                 1.18              16.05%                 5.15
## 2                 1.16              30.81%                 3.43
## 3                 0.98              31.63%                 1.93
## 4                 0.99              19.57%                 4.39
## 5                 1.09              22.34%                 2.64
## 6                 1.35              25.75%                 4.44
##   Total.Levered.Beta Current.PE Trailing.PE Forward.PE
## 1               7.36      72.98       30.35      27.52
## 2               3.76      29.78       31.15      30.91
## 3               3.09      47.14       28.11      14.42
## 4               5.07      27.86       27.82      23.93
## 5               4.90      13.55       15.06      29.57
## 6               5.24      20.77       20.35      16.57
##   Aggregate.Mkt.Cap..Net.Income Aggregate.Mkt.Cap..Trailing.Net.Income
## 1                         31.35                                  22.35
## 2                         19.52                                  17.81
## 3                         14.65                                  10.90
## 4                         30.23                                  24.26
## 5                         10.00                                  13.03
## 6                         18.68                                  17.42
##   Expected.growth...next.5.years PEG.Ratio  X
## 1                         13.08%      1.71 NA
## 2                         10.82%      1.65 NA
## 3                         34.73%      0.31 NA
## 4                         17.01%      1.43 NA
## 5                         21.93%      0.59 NA
## 6                         15.62%      1.11 NA
```

```
summary(raw_data)
```

```
##    Industry          Number.of.firms  Average.Unlevered.Beta
##  Length:96          Min.   :   4.0   Min.   :0.0600
##  Class :character   1st Qu.:  22.0   1st Qu.:0.7375
##  Mode  :character   Median :  46.5   Median :0.8950
##                     Mean   : 164.3   Mean   :0.8722
##                     3rd Qu.: 107.0   3rd Qu.:1.0025
##                     Max.   :7887.0   Max.   :1.4700
##
##  Average.Levered.Beta Average.correlation Total.Unlevered.Beta
##  Min.   :0.1000       Length:96           Min.   :0.310
##  1st Qu.:0.9975       Class :character    1st Qu.:2.420
##  Median :1.1050       Mode  :character    Median :3.300
##  Mean   :1.1101                           Mean   :3.626
##  3rd Qu.:1.2475                           3rd Qu.:4.647
##  Max.   :1.8200                           Max.   :9.470
##
##  Total.Levered.Beta   Current.PE       Trailing.PE       Forward.PE
##  Min.   : 1.350     Min.   :   9.22   Min.   :   9.47   Min.   : 10.92
##  1st Qu.: 3.312     1st Qu.:  24.55   1st Qu.:  22.59   1st Qu.: 19.48
##  Median : 4.260     Median :  37.83   Median :  30.75   Median : 24.76
##  Mean   : 4.584     Mean   :  83.54   Mean   :  59.96   Mean   : 32.78
##  3rd Qu.: 5.515     3rd Qu.:  86.01   3rd Qu.:  53.02   3rd Qu.: 34.24
##  Max.   :11.700     Max.   :1482.37   Max.   :1476.10   Max.   :261.70
##                                                         NA's   :1
```

```
##  Aggregate.Mkt.Cap..Net.Income Aggregate.Mkt.Cap..Trailing.Net.Income
##  Min.   :  2.93                Min.   :  6.51
##  1st Qu.: 17.32                1st Qu.: 14.38
##  Median : 22.25                Median : 20.09
##  Mean   : 30.63                Mean   : 20.85
##  3rd Qu.: 30.14                3rd Qu.: 23.97
##  Max.   :238.30                Max.   :126.57
##  NA's   :6
##  Expected.growth...next.5.years   PEG.Ratio         X
##  Length:96                       Min.   :0.170   Mode:logical
##  Class :character                1st Qu.:1.012   NA's:96
##  Mode  :character                Median :1.320
##                                  Mean   :1.498
##                                  3rd Qu.:1.657
##                                  Max.   :9.280
##                                  NA's   :2
```

```r
names(raw_data)
```

```
##  [1] "Industry"
##  [2] "Number.of.firms"
##  [3] "Average.Unlevered.Beta"
##  [4] "Average.Levered.Beta"
##  [5] "Average.correlation"
##  [6] "Total.Unlevered.Beta"
##  [7] "Total.Levered.Beta"
##  [8] "Current.PE"
##  [9] "Trailing.PE"
## [10] "Forward.PE"
## [11] "Aggregate.Mkt.Cap..Net.Income"
## [12] "Aggregate.Mkt.Cap..Trailing.Net.Income"
## [13] "Expected.growth...next.5.years"
## [14] "PEG.Ratio"
## [15] "X"
```

```r
str(raw_data)
```

```
## 'data.frame':    96 obs. of  15 variables:
##  $ Industry                              : chr  "Advertising" "Aerospace/Defense" "Air Transport" "Ap
##  $ Number.of.firms                       : int  52 93 22 64 22 75 13 676 22 46 ...
##  $ Average.Unlevered.Beta                : num  0.83 1.06 0.61 0.86 0.59 1.14 0.34 0.37 0.89 0.98 ..
##  $ Average.Levered.Beta                  : num  1.18 1.16 0.98 0.99 1.09 1.35 0.81 0.53 1.06 1.14 ..
##  $ Average.correlation                   : chr  "16.05%" "30.81%" "31.63%" "19.57%" ...
##  $ Total.Unlevered.Beta                  : num  5.15 3.43 1.93 4.39 2.64 4.44 0.79 1.68 4.83 6.16 ..
##  $ Total.Levered.Beta                    : num  7.36 3.76 3.09 5.07 4.9 5.24 1.88 2.37 5.71 7.18 ...
##  $ Current.PE                            : num  73 29.8 47.1 27.9 13.6 ...
##  $ Trailing.PE                           : num  30.4 31.1 28.1 27.8 15.1 ...
##  $ Forward.PE                            : num  27.5 30.9 14.4 23.9 29.6 ...
##  $ Aggregate.Mkt.Cap..Net.Income         : num  31.4 19.5 14.7 30.2 10 ...
##  $ Aggregate.Mkt.Cap..Trailing.Net.Income: num  22.4 17.8 10.9 24.3 13 ...
##  $ Expected.growth...next.5.years        : chr  "13.08%" "10.82%" "34.73%" "17.01%" ...
##  $ PEG.Ratio                             : num  1.71 1.65 0.31 1.43 0.59 1.11 1.69 1.38 1.4 1.89 ...
##  $ X                                     : logi  NA NA NA NA NA NA ...
```

```r
# Inspect individual elements of the merged data
print("Summary of elements in raw_data")
```

```
## [1] "Summary of elements in raw_data"
```

```r
for (i in 1:length(raw_data)) {
    print(paste0("Summary of ", names(raw_data)[i], ":"))
    print(summary(raw_data[, i]))
}
```

```
## [1] "Summary of Industry:"
##    Length     Class      Mode 
##        96 character character 
## [1] "Summary of Number.of.firms:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##     4.0    22.0    46.5   164.3   107.0  7887.0 
## [1] "Summary of Average.Unlevered.Beta:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.0600  0.7375  0.8950  0.8722  1.0020  1.4700 
## [1] "Summary of Average.Levered.Beta:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.1000  0.9975  1.1050  1.1100  1.2480  1.8200 
## [1] "Summary of Average.correlation:"
##    Length     Class      Mode 
##        96 character character 
## [1] "Summary of Total.Unlevered.Beta:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   0.310   2.420   3.300   3.626   4.648   9.470 
## [1] "Summary of Total.Levered.Beta:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   1.350   3.312   4.260   4.584   5.515  11.700 
## [1] "Summary of Current.PE:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    9.22   24.55   37.83   83.54   86.01 1482.00 
## [1] "Summary of Trailing.PE:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    9.47   22.59   30.75   59.96   53.02 1476.00 
## [1] "Summary of Forward.PE:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   10.92   19.48   24.76   32.78   34.25  261.70       1 
## [1] "Summary of Aggregate.Mkt.Cap..Net.Income:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##    2.93   17.31   22.25   30.63   30.14  238.30       6 
## [1] "Summary of Aggregate.Mkt.Cap..Trailing.Net.Income:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    6.51   14.38   20.09   20.85   23.97  126.60 
## [1] "Summary of Expected.growth...next.5.years:"
##    Length     Class      Mode 
##        96 character character 
## [1] "Summary of PEG.Ratio:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   0.170   1.013   1.320   1.498   1.657   9.280       2 
## [1] "Summary of X:"
```

```
##    Mode    NA's
## logical      96
```

Further inspection revealed that column "X" is a column full of NA values We start to clean data by removing this column through subsetting and reassigning this subset as the dataframe "clean_data".

```r
# Remove blank columns
clean_data <- subset(raw_data, select = -X)
```

We also find that some columns have unideal names, which we rename.

```r
# Rename column titled Aggregate.Mkt.Cap..Net.Income
clean_data$Aggregate.Mkt.Cap.Net.Income <- clean_data$Aggregate.Mkt.Cap..Net.Income
clean_data$Aggregate.Mkt.Cap..Net.Income <- NULL

# Rename column titled clean_data$Aggregate.Mkt.Cap..Trailing.Net.Income
clean_data$Aggregate.Mkt.Cap.Trailing.Net.Income <- clean_data$Aggregate.Mkt.Cap..Trailing.Net.Income
clean_data$Aggregate.Mkt.Cap..Trailing.Net.Income <- NULL

# Rename column titled Number.of.firms
clean_data$Number.of.Firms <- clean_data$Number.of.firms
clean_data$Number.of.firms <- NULL
```

We then turn the two character columns, Average Correlation and Expected Growth in Next 5 Years, into numeric vectors for more accurate analysis of numbers.

```r
# Rename column titled Average.correlation and turn it into a numeric column
# vector
clean_data$Average.Correlation <- as.numeric(gsub("%", "", clean_data$Average.correlation))
clean_data$Average.correlation <- NULL
head(clean_data$Average.Correlation)
```

```
## [1] 16.05 30.81 31.63 19.57 22.34 25.75
```

```r
# Rename column titled Expected.growth...next.5.years and turn it into a
# numeric column vector
clean_data$Expected.Growth.Next.5.Years <- as.numeric(gsub("%", "", clean_data$Expected.growth...next.5
clean_data$Expected.growth...next.5.years <- NULL
head(clean_data$Expected.Growth.Next.5.Years)
```

```
## [1] 13.08 10.82 34.73 17.01 21.93 15.62
```

Then, we proceed to removing rows that contain data that cannot be evaluated or will not aid in our analysis. We found that the row "Unclassified" would fit that description. We also considered removing "Total Market", since that did not pertain to the entire industry but is simply an aggregate of all other industries.

```r
# Make copy of clean_data without the row 'Unclassified'
clean_data <- clean_data[-which(clean_data$Industry == "Unclassified"), ]

# Make copy of clean_data without the row 'Unclassified' or 'Total Market'
industries_only <- clean_data[-which(clean_data$Industry == "Unclassified" |
    clean_data$Industry == "Total Market"), ]
```

We then inspect elements of our cleaned datasets.

```r
# Inspect elements of the final clean_data and industries_only
print("Summary of elements in clean_data")
```

```
## [1] "Summary of elements in clean_data"
```

```r
for (i in 1:length(clean_data)) {
    print(paste0("Summary of ", names(clean_data)[i], ":"))
    print(summary(clean_data[, i]))
}
```

```
## [1] "Summary of Industry:"
##    Length     Class      Mode
##        95 character character
## [1] "Summary of Average.Unlevered.Beta:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0600  0.7450  0.9000  0.8799  1.0050  1.4700
## [1] "Summary of Average.Levered.Beta:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.530   1.010   1.110   1.121   1.255   1.820
## [1] "Summary of Total.Unlevered.Beta:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.310   2.450   3.310   3.640   4.685   9.470
## [1] "Summary of Total.Levered.Beta:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.350   3.330   4.320   4.615   5.520  11.700
## [1] "Summary of Current.PE:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.22   24.80   38.02   84.21   86.34 1482.00
## [1] "Summary of Trailing.PE:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.47   22.82   31.15   60.39   53.14 1476.00
## [1] "Summary of Forward.PE:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.92   19.48   24.76   32.78   34.25  261.70
## [1] "Summary of PEG.Ratio:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.170   1.013   1.320   1.498   1.657   9.280       1
## [1] "Summary of Aggregate.Mkt.Cap.Net.Income:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    2.93   17.31   22.25   30.63   30.14  238.30       5
## [1] "Summary of Aggregate.Mkt.Cap.Trailing.Net.Income:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    6.51   14.47   20.18   20.94   24.02  126.60
## [1] "Summary of Number.of.Firms:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     4.0    22.5    47.0   166.0   111.0  7887.0
## [1] "Summary of Average.Correlation:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   11.05   20.45   25.75   27.32   32.16   51.37
## [1] "Summary of Expected.Growth.Next.5.Years:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -23.28   11.76   14.45   15.02   18.12   37.65
```

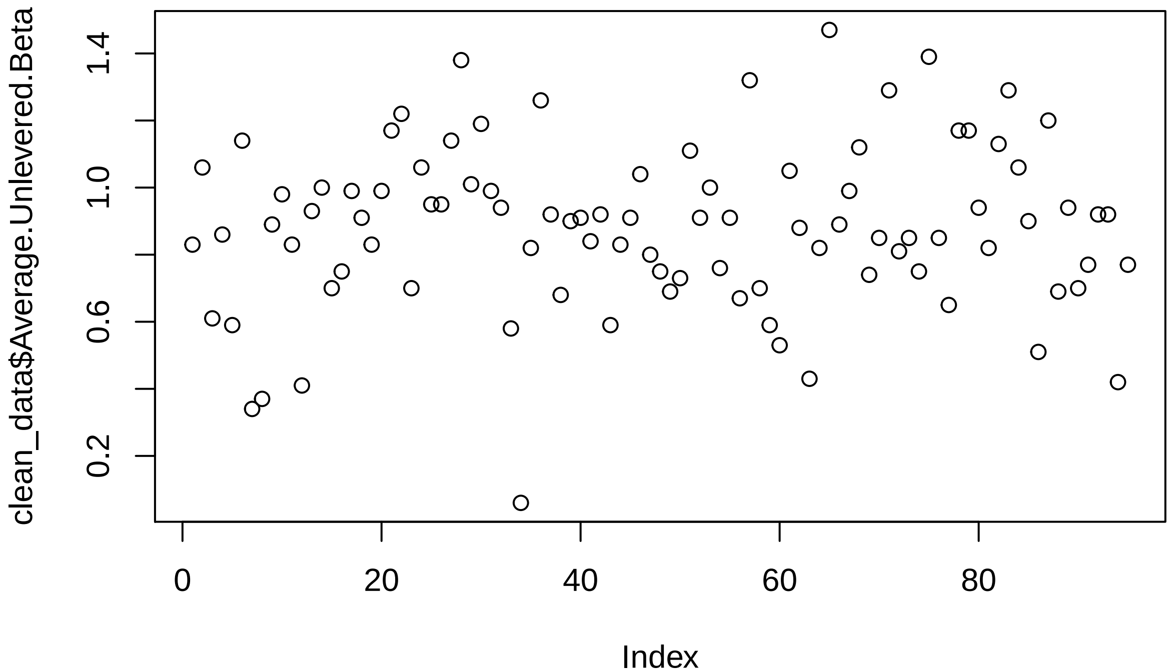
```r
print("Summary of elements in industries_only")
```

```
## [1] "Summary of elements in industries_only"
```

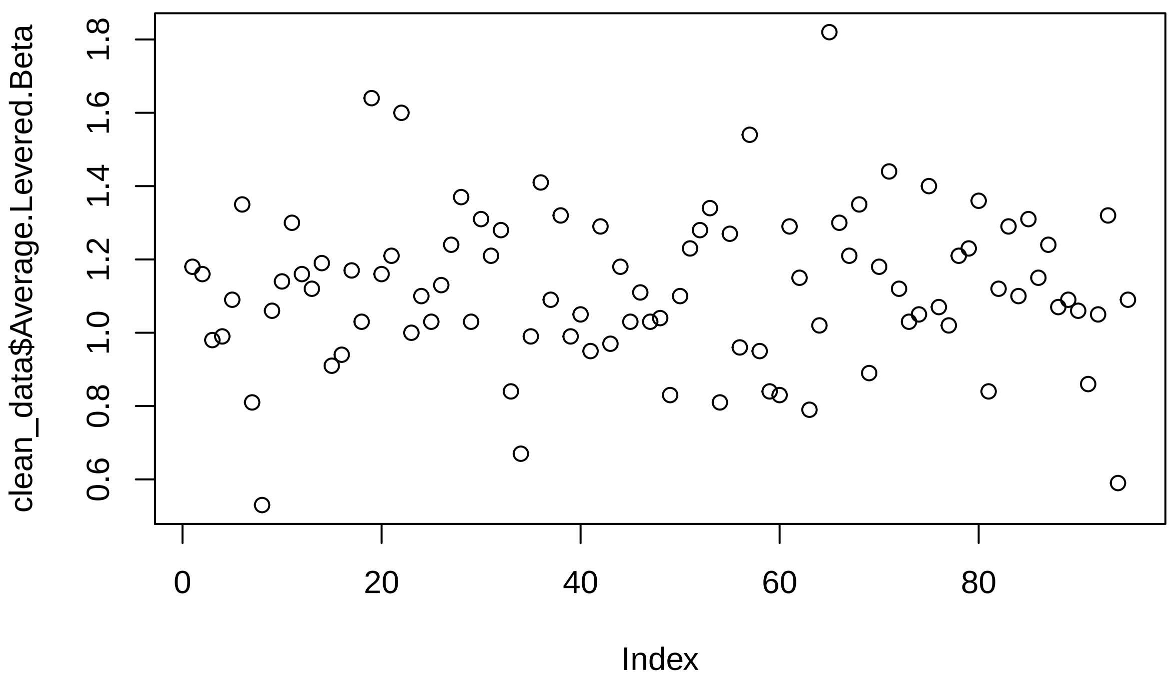
```r
for (i in 1:length(industries_only)) {
    print(paste0("Summary of ", names(industries_only)[i], ":"))
    print(summary(industries_only[, i]))
}
```

```
## [1] "Summary of Industry:"
##    Length     Class      Mode 
##        94 character character 
## [1] "Summary of Average.Unlevered.Beta:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.0600  0.7500  0.9000  0.8818  1.0080  1.4700 
## [1] "Summary of Average.Levered.Beta:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   0.530   1.005   1.115   1.121   1.262   1.820 
## [1] "Summary of Total.Unlevered.Beta:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   0.310   2.440   3.310   3.648   4.722   9.470 
## [1] "Summary of Total.Levered.Beta:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   1.350   3.325   4.260   4.616   5.525  11.700 
## [1] "Summary of Current.PE:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    9.22   24.69   37.83   84.34   86.68 1482.00 
## [1] "Summary of Trailing.PE:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    9.47   22.80   30.75   60.39   52.67 1476.00 
## [1] "Summary of Forward.PE:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   10.92   19.46   24.66   32.75   33.26  261.70 
## [1] "Summary of PEG.Ratio:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   0.170   1.010   1.330   1.502   1.660   9.280       1 
## [1] "Summary of Aggregate.Mkt.Cap.Net.Income:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##    2.93   17.29   22.35   30.76   30.23  238.30       5 
## [1] "Summary of Aggregate.Mkt.Cap.Trailing.Net.Income:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    6.51   14.44   20.25   20.97   24.06  126.60 
## [1] "Summary of Number.of.Firms:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    4.00   22.25   46.50   83.82  103.00  676.00 
## [1] "Summary of Average.Correlation:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   11.05   20.39   25.78   27.36   32.32   51.37 
## [1] "Summary of Expected.Growth.Next.5.Years:"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  -23.28   11.75   14.44   15.02   18.17   37.65
```

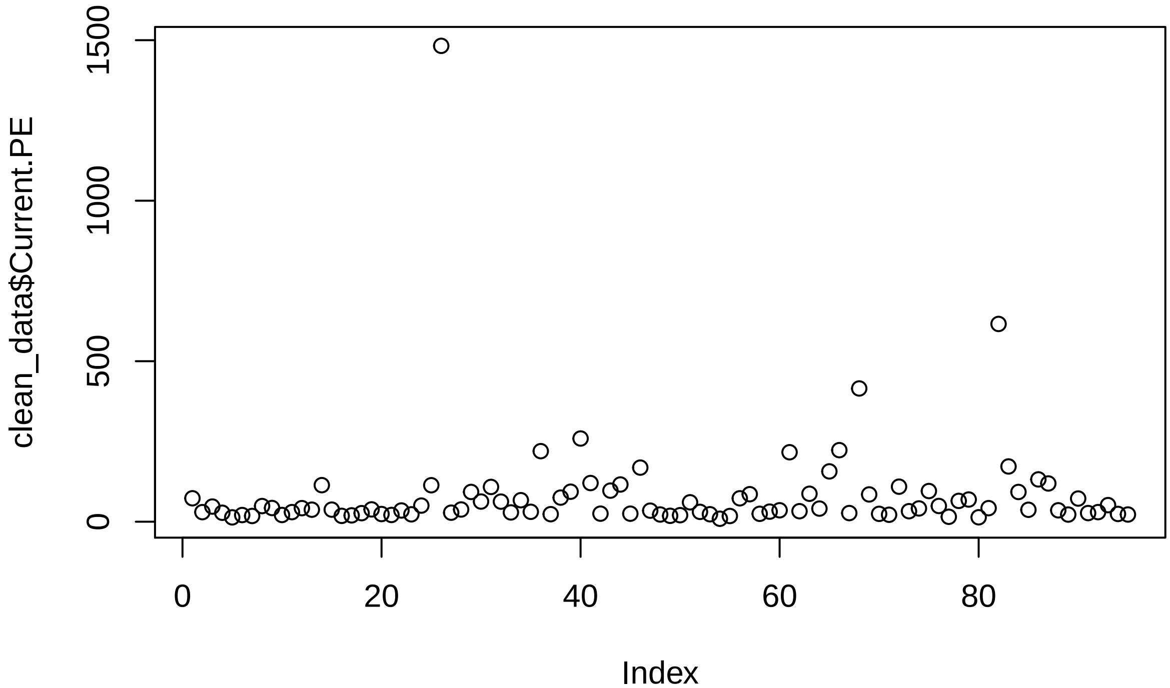
```r
# Basic plots of variables
plot(clean_data$Average.Unlevered.Beta)
```



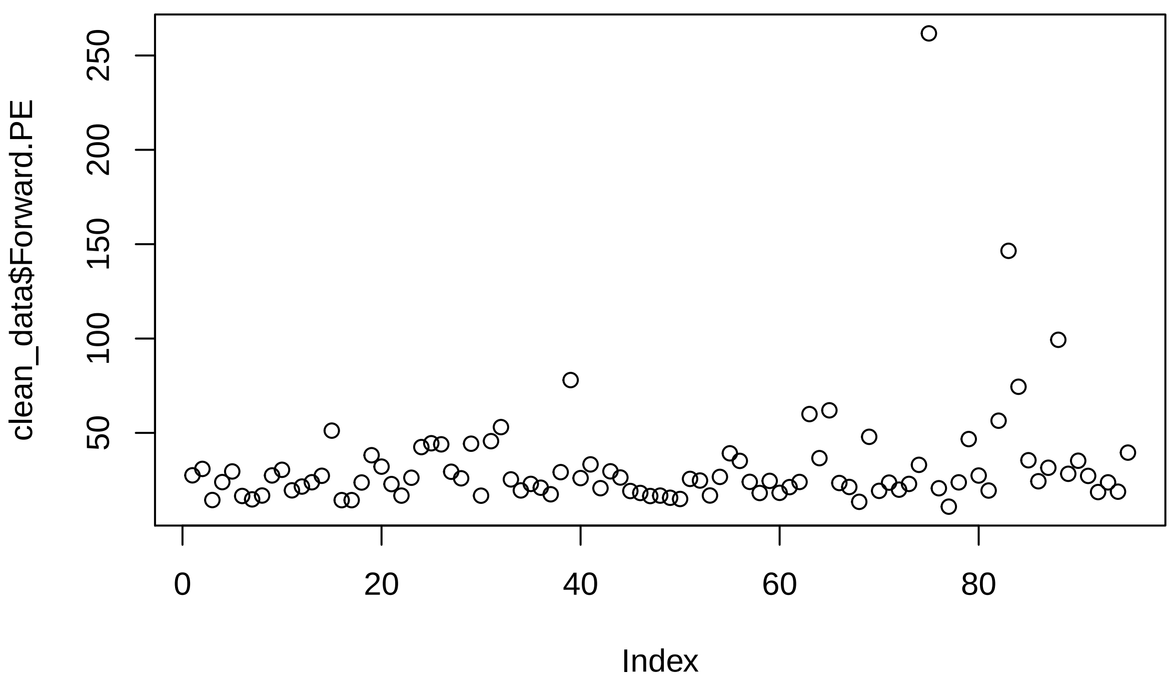

```r
plot(clean_data$Average.Levered.Beta)
```



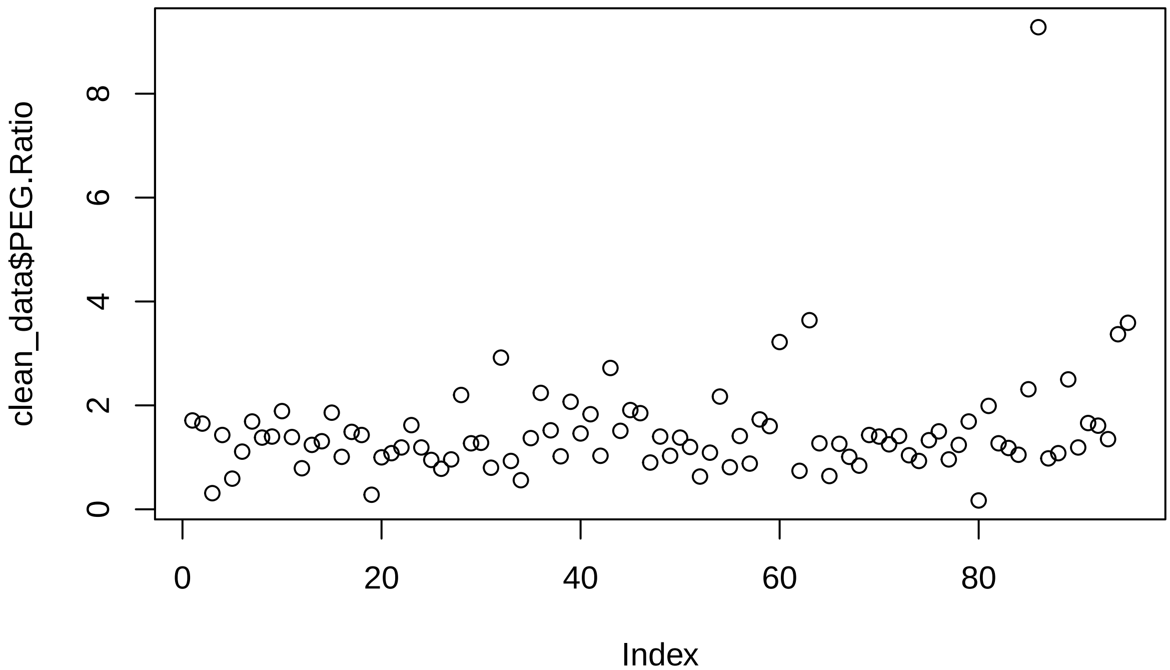

```r
plot(clean_data$Current.PE)
```



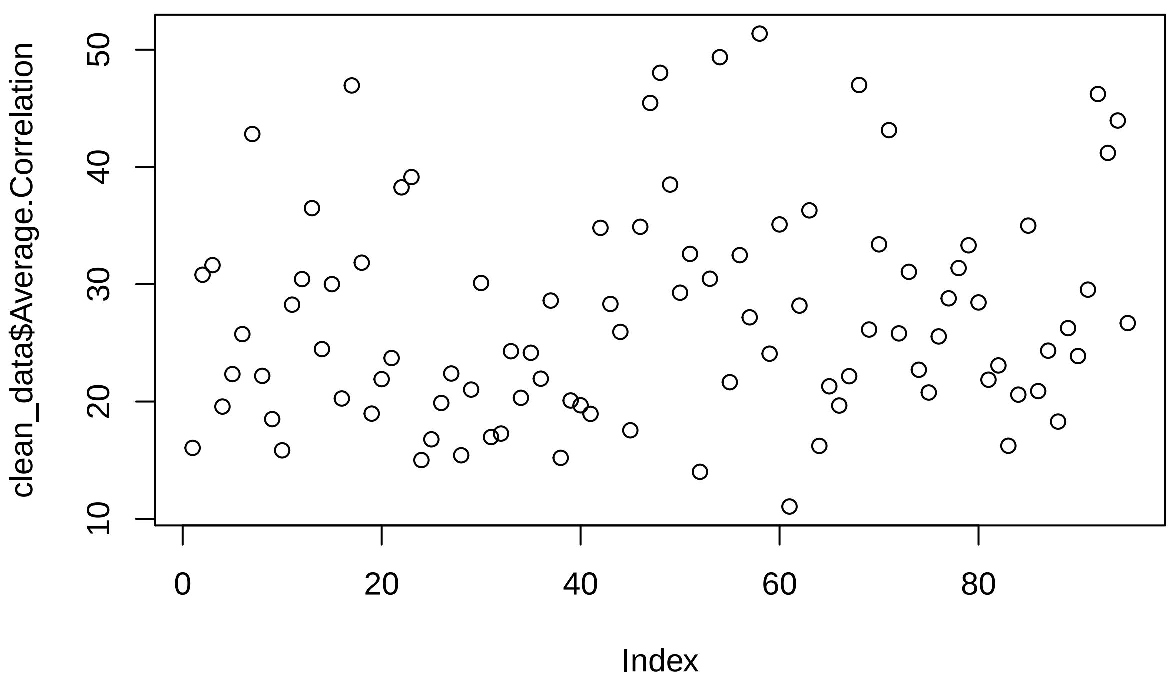

```r
plot(clean_data$Forward.PE)
```



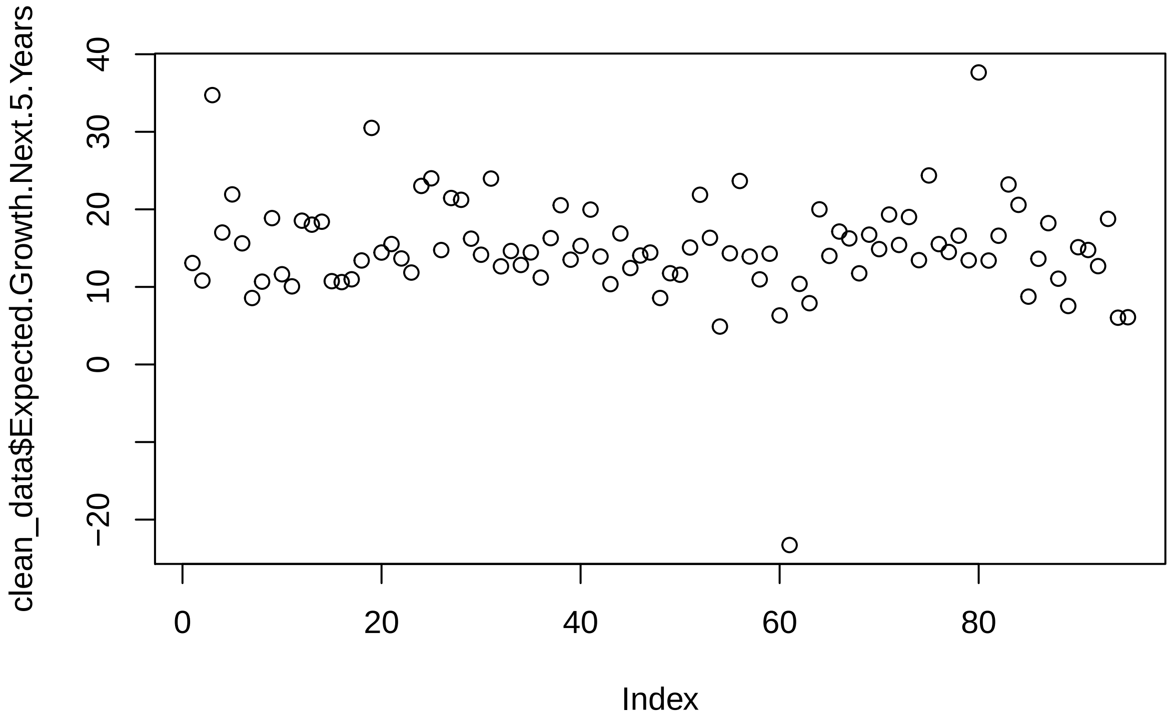

```r
plot(clean_data$PEG.Ratio)
```




```
plot(clean_data$Average.Correlation)
```




```
plot(clean_data$Expected.Growth.Next.5.Years)
```




```r
# Exploration of industries that have certain maximum and minimum values
clean_data[which.max(clean_data$Average.Unlevered.Beta), "Industry"]
```

```
## [1] "Real Estate (General/Diversified)"
```

```r
clean_data[which.min(clean_data$Average.Unlevered.Beta), "Industry"]
```

```
## [1] "Financial Svcs. (Non-bank & Insurance)"
```

```r
clean_data[which.max(clean_data$Average.Levered.Beta), "Industry"]
```

```
## [1] "Real Estate (General/Diversified)"
```

```r
clean_data[which.min(clean_data$Average.Levered.Beta), "Industry"]
```

```
## [1] "Banks (Regional)"
```

```r
clean_data[which.max(clean_data$Current.PE), "Industry"]
```

```
## [1] "Education"
```

```r
clean_data[which.min(clean_data$Current.PE), "Industry"]
```

```
## [1] "Oil/Gas (Integrated)"
```

```r
clean_data[which.max(clean_data$PEG.Ratio), "Industry"]
```

```
## [1] "Telecom (Wireless)"
```

```r
clean_data[which.min(clean_data$PEG.Ratio), "Industry"]
```

```
## [1] "Shipbuilding & Marine"
```

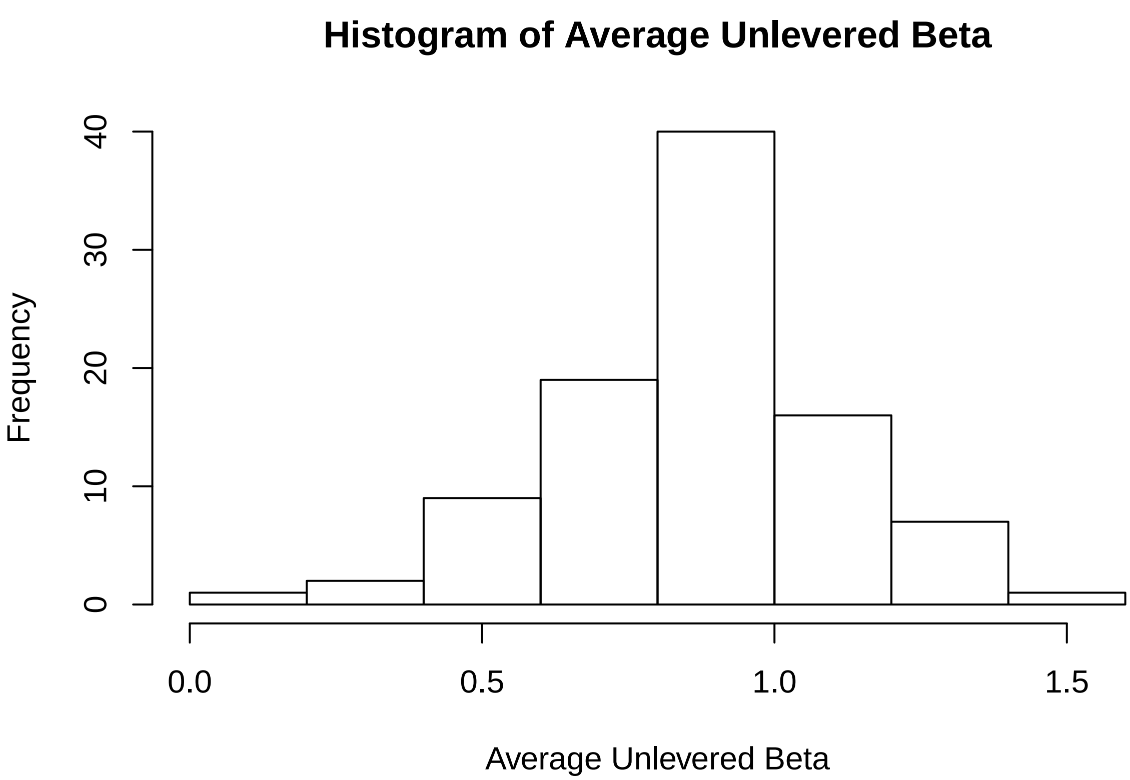
```r
clean_data[which.max(clean_data$Expected.Growth.Next.5.Years), "Industry"]
```

```
## [1] "Shipbuilding & Marine"
```

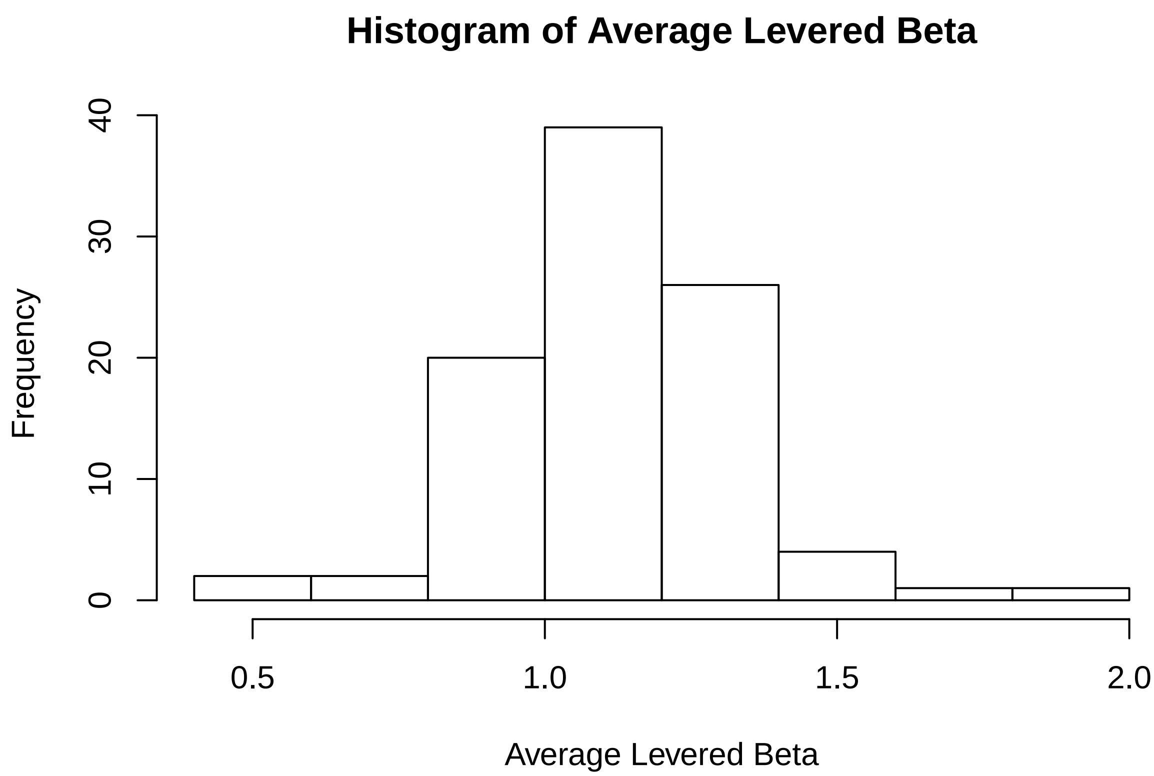
```r
clean_data[which.min(clean_data$Expected.Growth.Next.5.Years), "Industry"]
```

```
## [1] "Precious Metals"
```

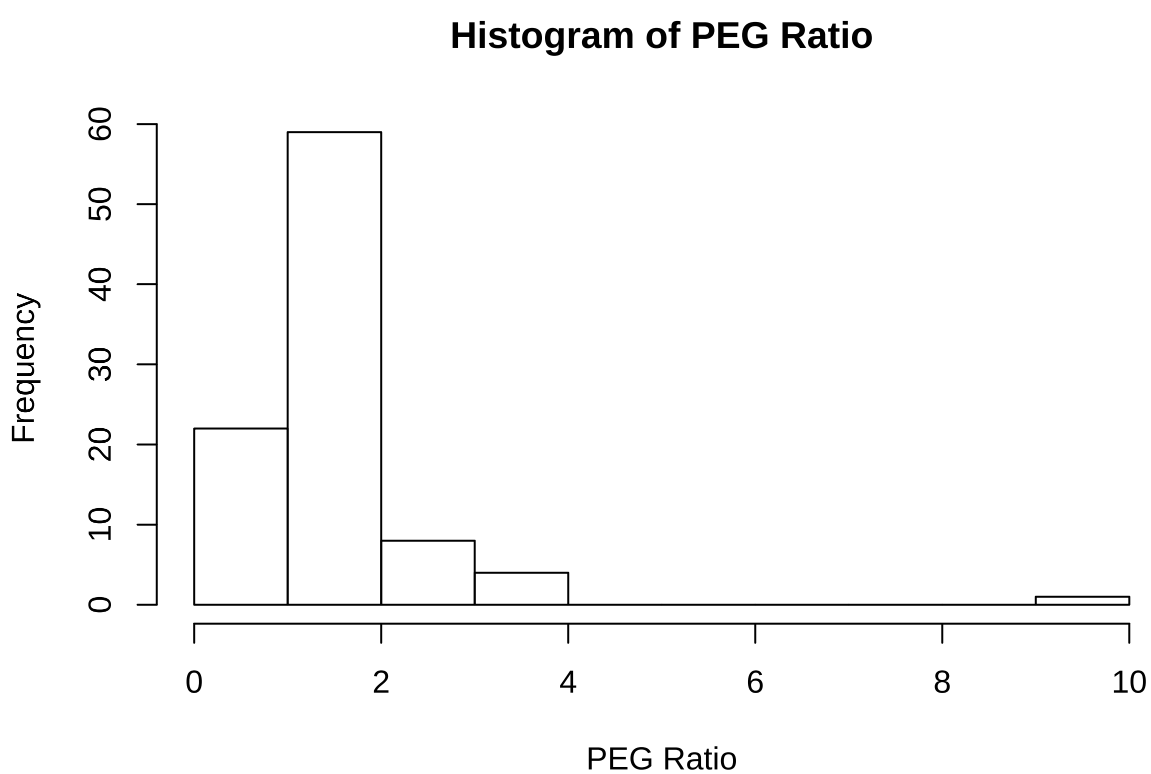
```r
# Histograms of select variables
hist(clean_data$Average.Unlevered.Beta, main = "Histogram of Average Unlevered Beta",
    xlab = "Average Unlevered Beta")
```



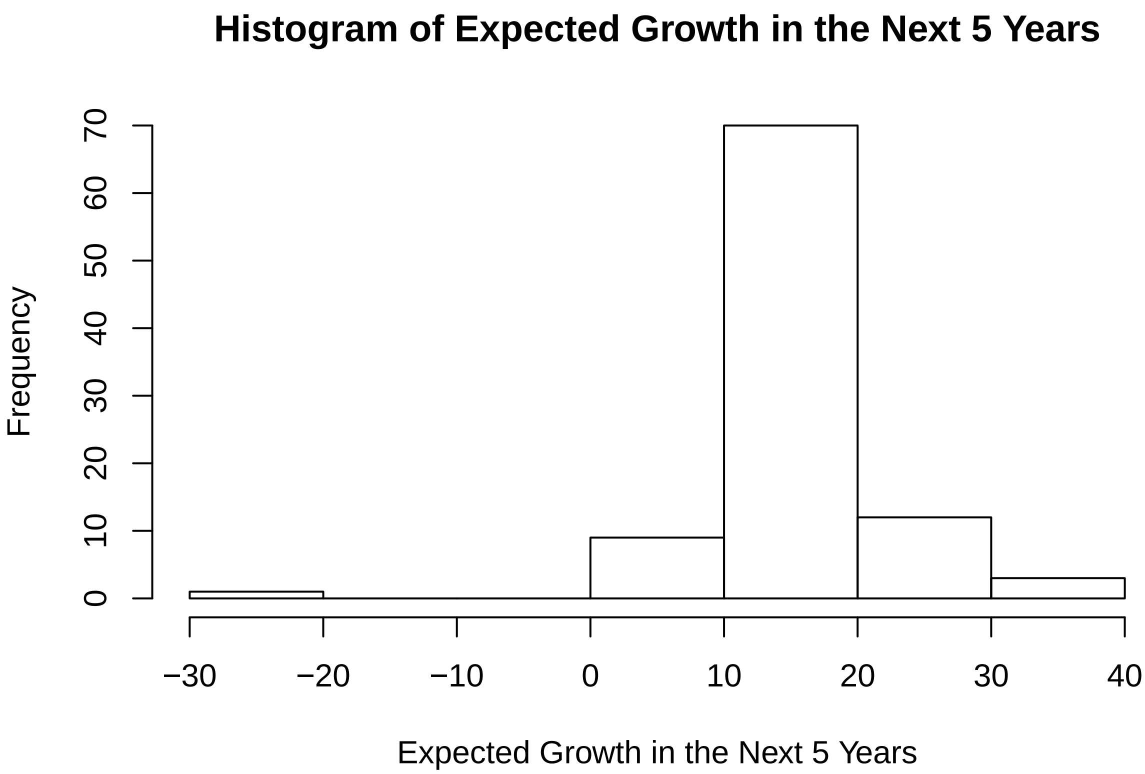

```r
hist(clean_data$Average.Levered.Beta, main = "Histogram of Average Levered Beta",
    xlab = "Average Levered Beta")
```




```r
hist(clean_data$PEG.Ratio, main = "Histogram of PEG Ratio", xlab = "PEG Ratio")
```



```r
hist(clean_data$Expected.Growth.Next.5.Years, main = "Histogram of Expected Growth in the Next 5 Years"
    xlab = "Expected Growth in the Next 5 Years")
```



Finally, we create CSV files for our cleaned data and placed it into the clean_data directory.

```r
# Create CSV files for clean data
file.create("clean_data/clean_data.csv")
write.csv(clean_data, file = "clean_data/clean_data.csv")
file.create("clean_data/industries_only.csv")
write.csv(industries_only, file = "clean_data/industries_only.csv")
```

Methods & Analysis

Before we can analyze our data, we will first need to set up the packages and dependencies we will be using. We will also need to retrieve and read our clean dataset.

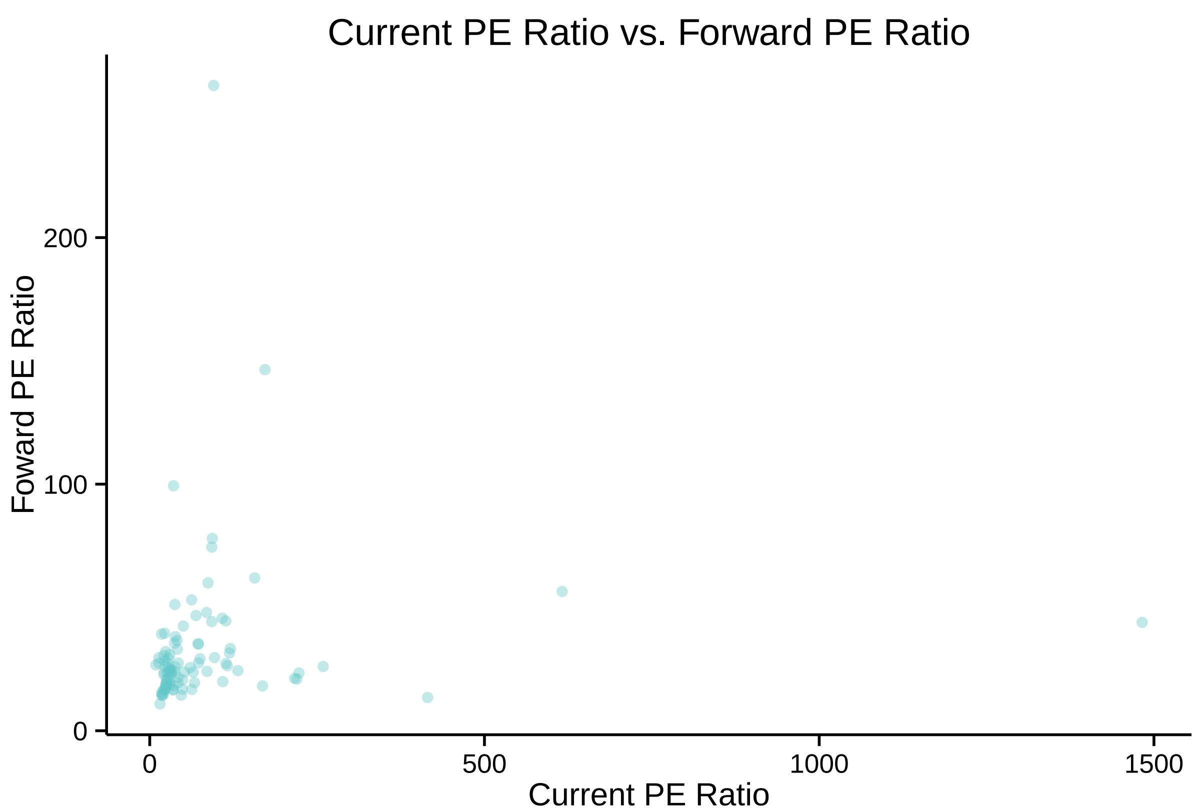
```r
# Set working directory back for knitting
setwd("/Users/Michelle/Documents/UC Berkeley 2015-2016/Statistics 133/projects/final/report")

# Set up ggplot2
library(ggplot2)
dev.off()
```

```
## null device
##           1
```

```r
# Set up readr
library(readr)

# Set up scatterplot3d
library(scatterplot3d)

# Set up stringr
library(stringr)

# Set correct working directory again
setwd("~/Documents/UC Berkeley 2015-2016/Statistics 133/projects/final/")

# Read data files
clean_data <- read.csv("clean_data/clean_data.csv", header = TRUE, stringsAsFactors = FALSE)
industries_only <- read.csv("clean_data/industries_only.csv", header = TRUE,
    stringsAsFactors = FALSE)
```

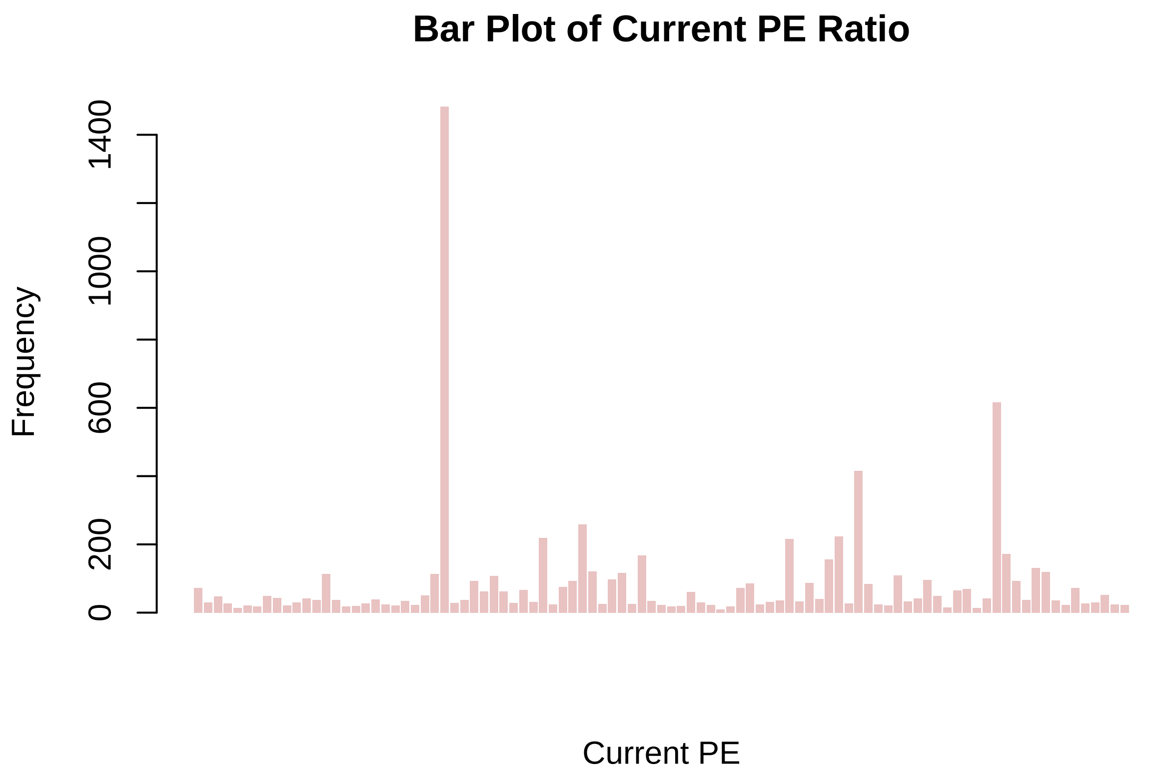
—— *ELLEN* ————

```r
# Scatter plot of Current PE to Forward PE
ggplot(clean_data, aes(x = Current.PE, y = Forward.PE)) + geom_point(color = rgb(100,
    200, 200, 100, maxColorValue = 255)) + ggtitle("Current PE Ratio vs. Forward PE Ratio") +
    xlab("Current PE Ratio") + ylab("Foward PE Ratio") + theme_classic()
```

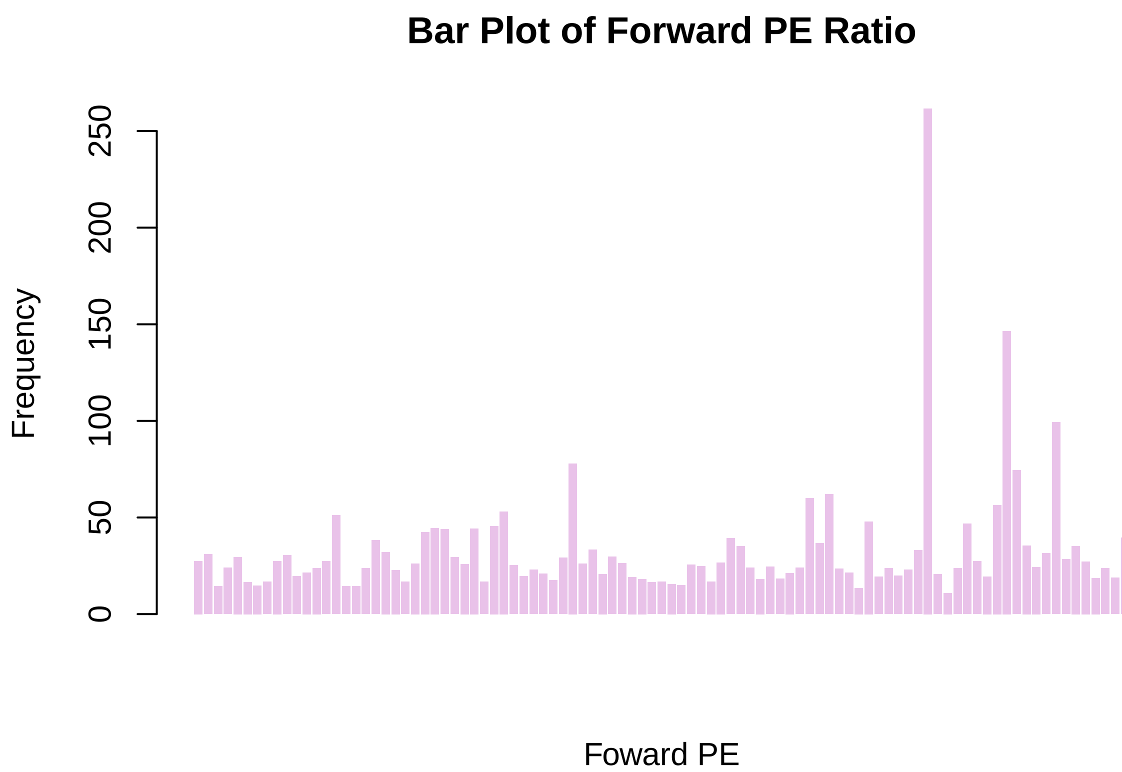



```r
# Bar plot of Current PE, Forward PE, Total Levered Beta, and Expected
# Growth in the Next 5 Years
```

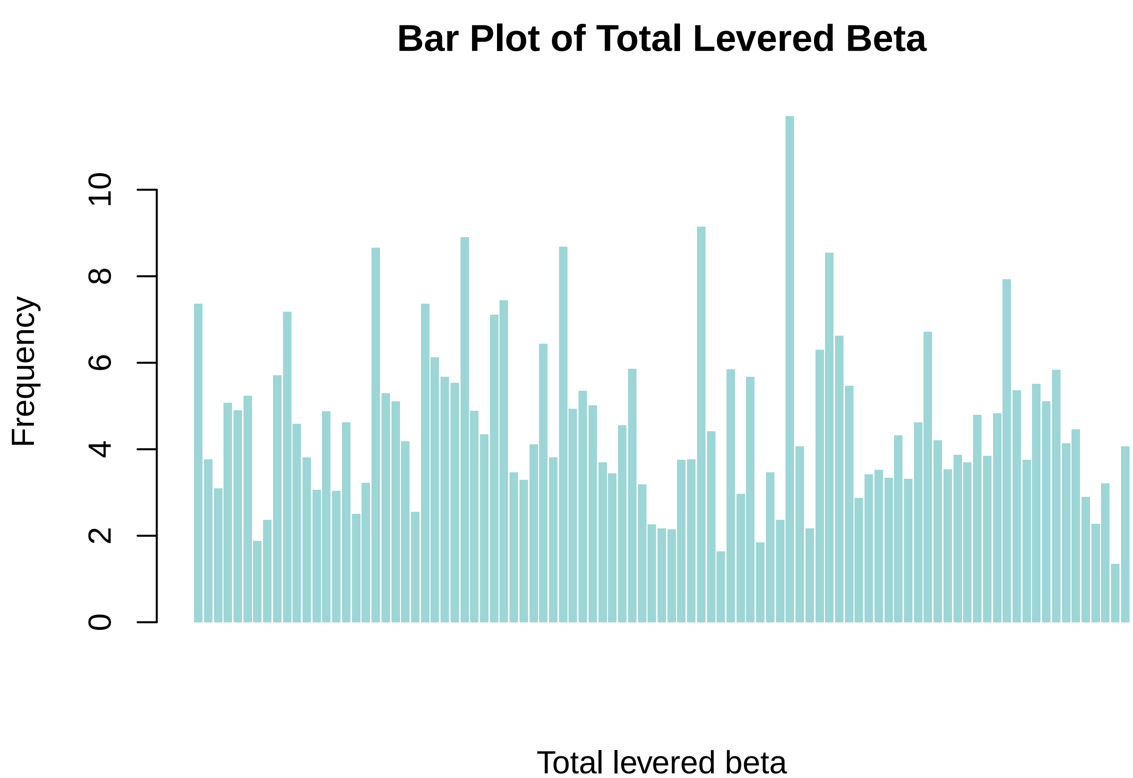
```
barplot(clean_data$Current.PE, border = NA, xlab = "Current PE", ylab = "Frequency",
    col = rgb(200, 100, 100, 100, maxColorValue = 255), main = "Bar Plot of Current PE Ratio")
```



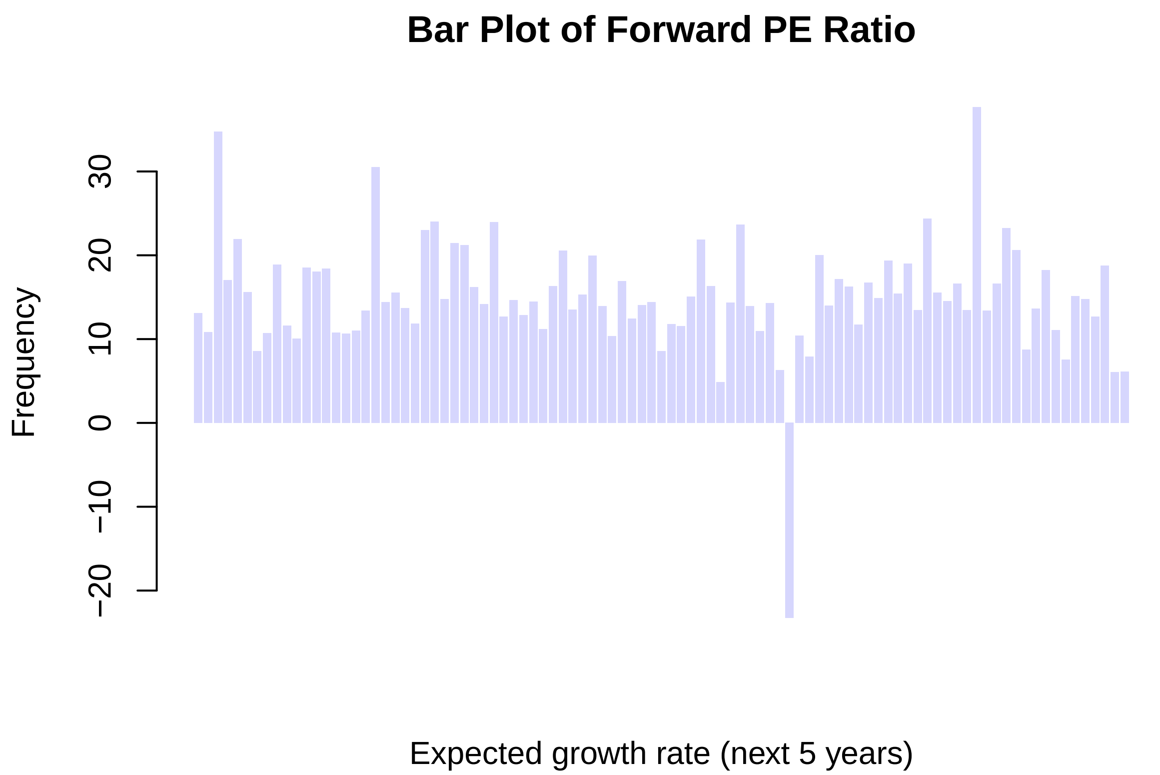

```
barplot(clean_data$Forward.PE, border = NA, xlab = "Foward PE", ylab = "Frequency",
    col = rgb(200, 100, 200, 100, maxColorValue = 255), main = "Bar Plot of Forward PE Ratio")
```




```
barplot(clean_data$Total.Levered.Beta, border = NA, xlab = "Total levered beta",
    ylab = "Frequency", col = rgb(0, 150, 150, 100, maxColorValue = 255), main = "Bar Plot of Total Lev
```



```r
barplot(clean_data$Expected.Growth.Next.5.Years, border = NA, xlab = "Expected growth rate (next 5 year
    ylab = "Frequency", col = rgb(150, 150, 250, 100, maxColorValue = 255),
    main = "Bar Plot of Forward PE Ratio")
```




```r
# Analyze which industries are in the highest and lowest beta interval
beta_intervals <- seq(from = 0, to = 1.5, by = 0.3)
interval_fun <- function(beta) {
    for (i in 1:5) {
        if (beta_intervals[i] < beta & beta < beta_intervals[i + 1]) {
            return(i)
        }
    }
}
beta_interval <- unlist(lapply(clean_data$Average.Unlevered.Beta, FUN = interval_fun))
as.vector(clean_data$Industry[unlist(beta_interval) == 5])
```

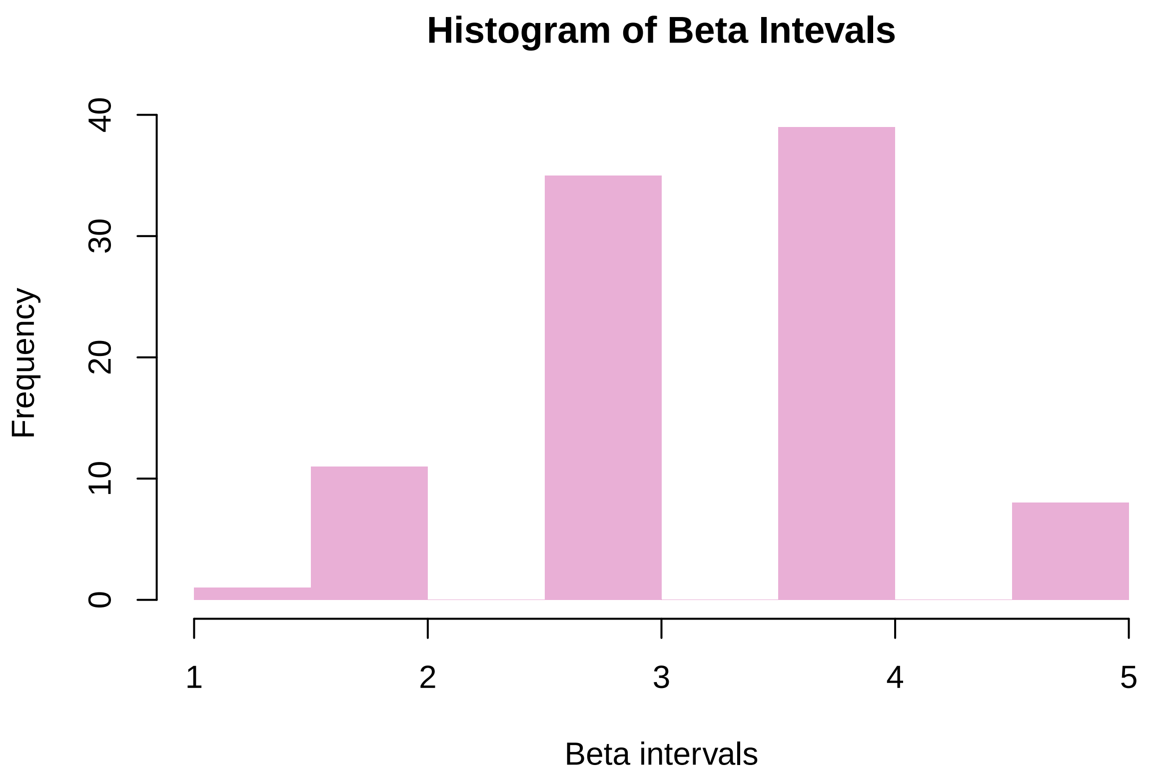
```
## [1] "Construction Supplies"              "Electronics (Consumer & Office)"
## [3] "Food Wholesalers"                   "Oilfield Svcs/Equip."
## [5] "Real Estate (General/Diversified)" "Retail (Building Supply)"
## [7] "Retail (Online)"                    "Software (Internet)"
```

```r
as.vector(clean_data$Industry[unlist(beta_interval) == 1])
```

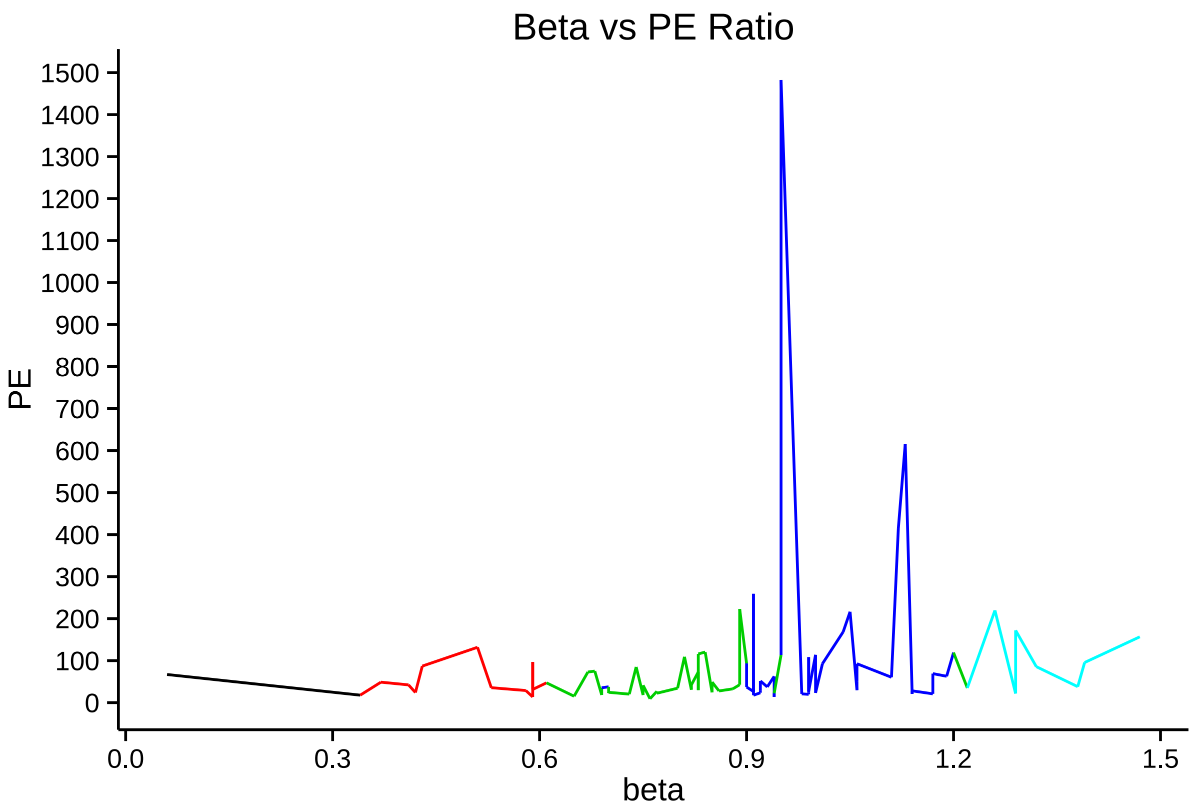
```
## [1] "Financial Svcs. (Non-bank & Insurance)"
```

We find that we have 8 industries in the highest beta interval, which are companies in growing markets such as online retail, real estate, etc. Companies with lowest beta are stable markets such as financial services (Non-Bank or insurance), or trucking.

```r
# Make a histogram of the beta intervals created above
dfbeta_interval <- data.frame(table(unlist(beta_interval)))
names(dfbeta_interval) <- c("Interval", "Frequency")
hist(beta_interval, main = "Histogram of Beta Intevals", xlab = "Beta intervals",
    col = rgb(200, 50, 150, 100, maxColorValue = 255), border = NA)
```




```r
# Plot of Beta versus PE Ratio by given beta intervals
dat1 <- data.frame(PE = industries_only$Current.PE, beta = industries_only$Average.Unlevered.Beta)
ggplot(dat1, aes(x = beta, y = PE)) + geom_line(stat = "identity", color = beta_interval) +
    ggtitle("Beta vs PE Ratio") + theme_classic() + scale_x_continuous(breaks = beta_intervals) +
    scale_y_continuous(breaks = seq(from = 0, to = 1500, by = 100))
```

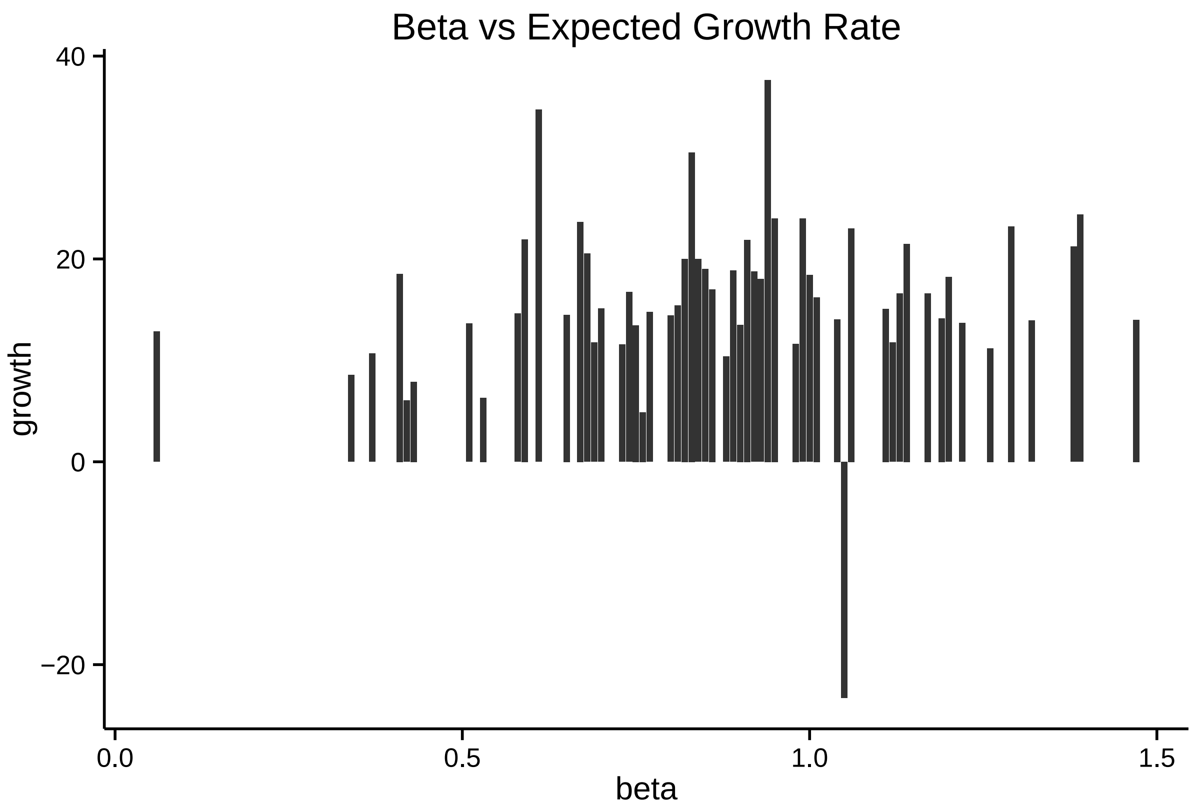



```r
summary(industries_only$Current.PE)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.22   24.69   37.83   84.34   86.68 1482.00
```

Industries that have average betas of 0.9-1.2 (Interval 4) have higher PE ratios.

```r
# Bar plot of Beta versus Expected Growth Rate in the Next 5 Years
dat2 <- data.frame(beta = clean_data$Average.Unlevered.Beta, growth = clean_data$Expected.Growth.Next.5
ggplot(dat2, aes(x = beta, y = growth)) + geom_bar(position = "identity", stat = "identity") +
    ggtitle("Beta vs Expected Growth Rate") + theme_classic()
```

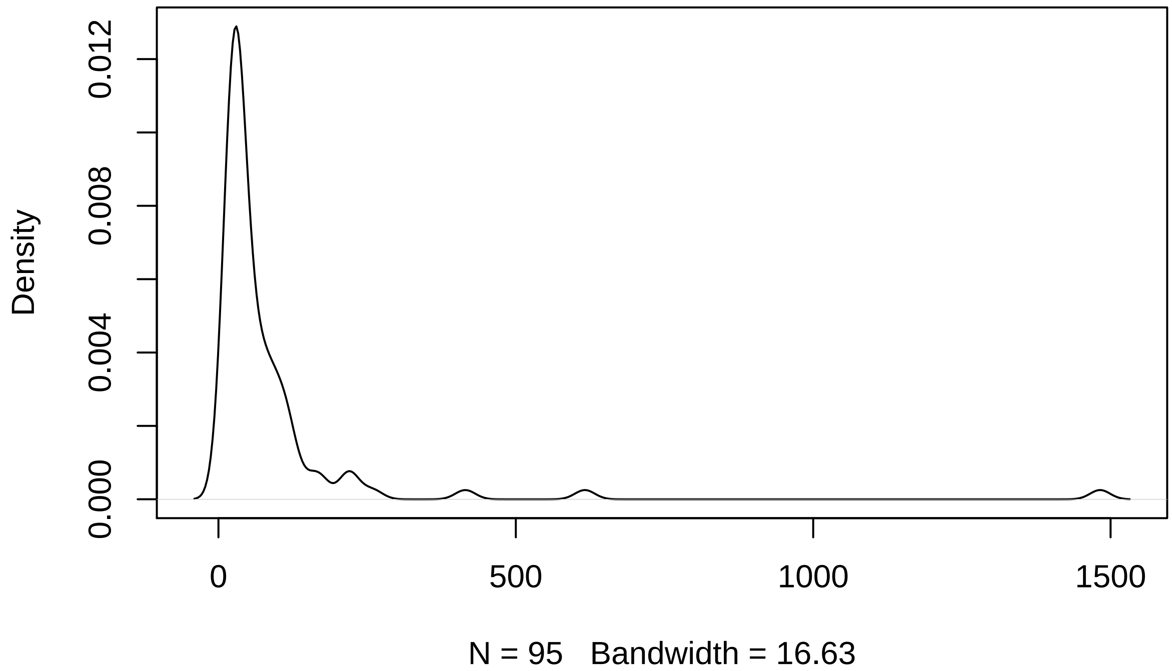


Industries with betas of around 1 have highter expected growth.

———— *Michelle* —————-

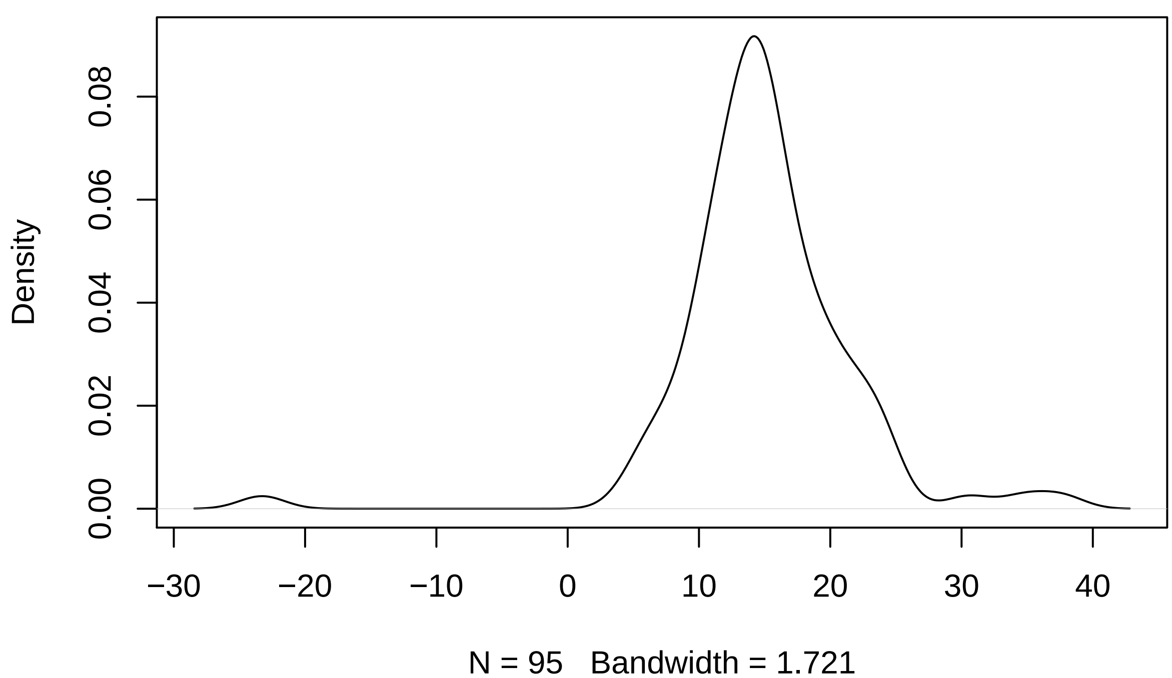
```
# Density curve of Current PE and of Expected Growth in the Next 5 Years
plot(density(clean_data$Current.PE), main = "Density Plot of Current PE")
```

Density Plot of Current PE

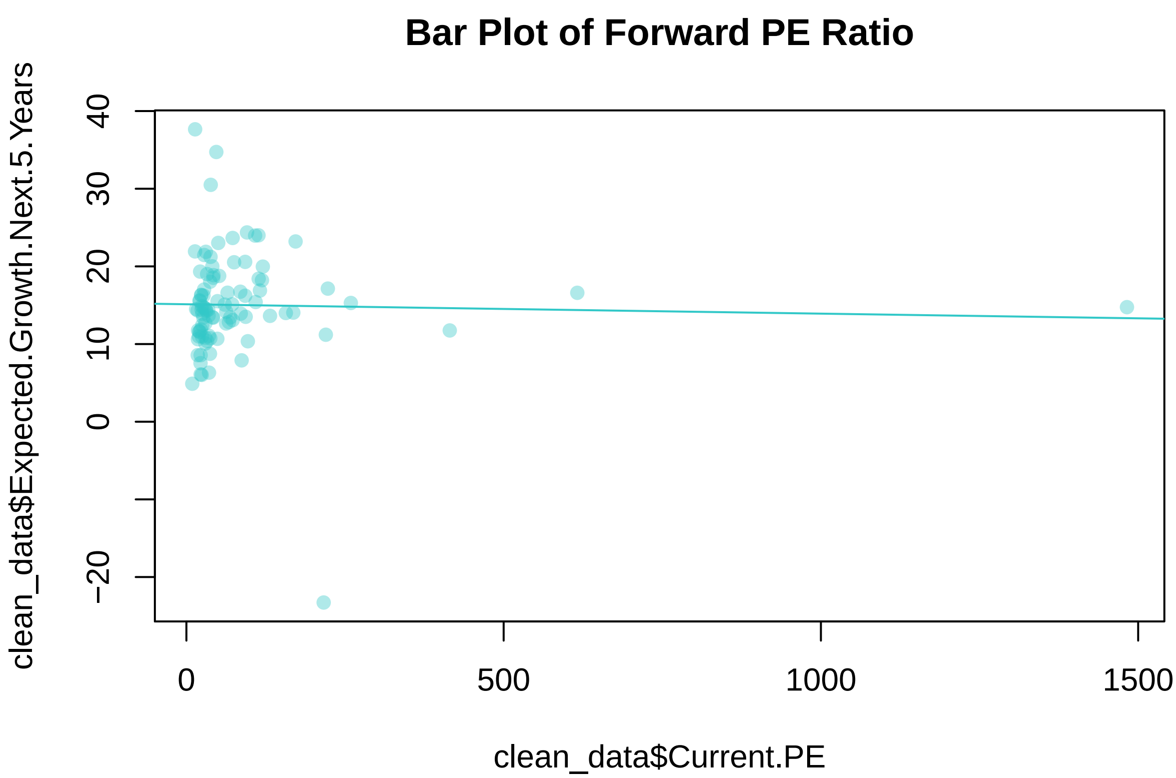



```r
plot(density(clean_data$Expected.Growth.Next.5.Years), main = "Density Plot of Expected Growth in the N
```

Density Plot of Expected Growth in the Next 5 Years



```r
# Scatter plot of Current PE to Expected Growth in the Next 5 Years
plot(clean_data$Current.PE, clean_data$Expected.Growth.Next.5.Years, col = rgb(50,
    200, 200, 100, maxColorValue = 255), pch = 16, bg = rgb(200, 200, 200, maxColorValue = 255),
    main = "Bar Plot of Forward PE Ratio")

# Make a linear regression of Current PE to Expected Growth in the Next 5
# Years
fit1 <- lm(Expected.Growth.Next.5.Years ~ Current.PE, data = clean_data)
abline(fit1, col = rgb(50, 200, 200, maxColorValue = 255))
```

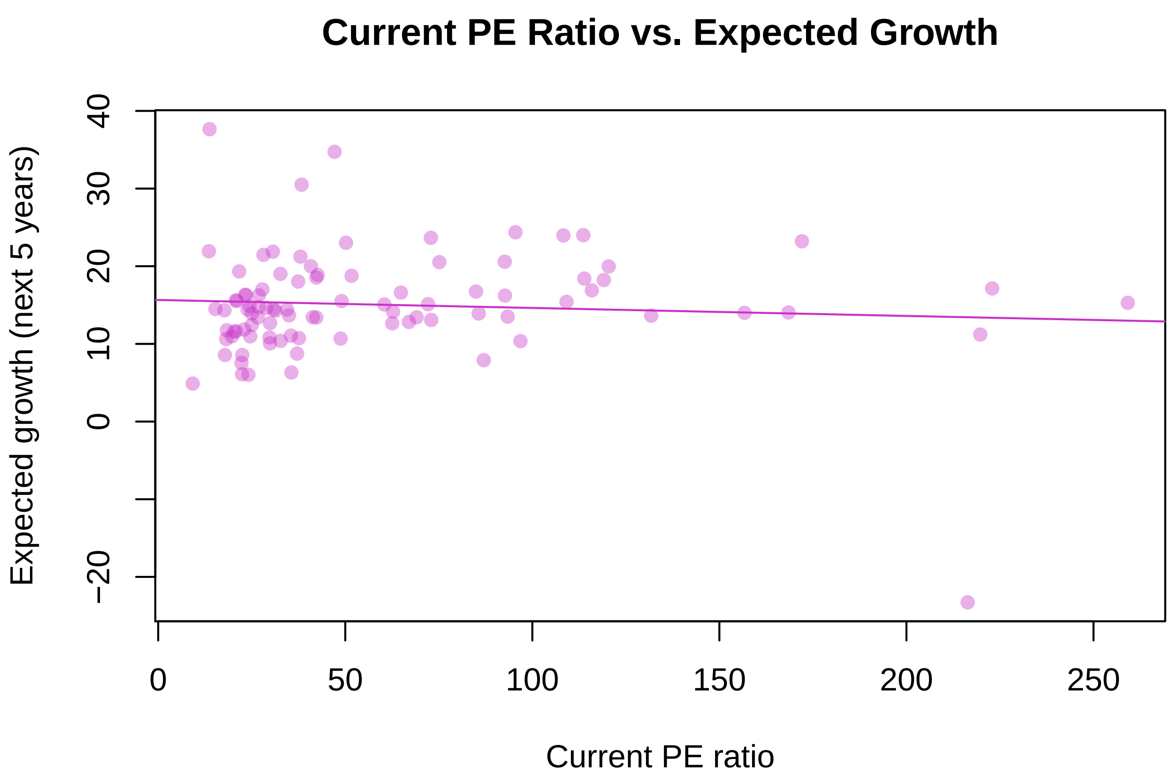



```r
summary(fit1)
```

```
##
## Call:
## lm(formula = Expected.Growth.Next.5.Years ~ Current.PE, data = clean_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -38.141  -3.293  -0.635   3.097  22.544
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15.122630   0.786276  19.233   <2e-16 ***
## Current.PE  -0.001207   0.004205  -0.287    0.775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.843 on 93 degrees of freedom
```

```
## Multiple R-squared:  0.000886,	Adjusted R-squared:  -0.009857
## F-statistic: 0.08247 on 1 and 93 DF,  p-value: 0.7746
```

```
# Remove outliers and find the linear regression of Current PE to Expected
# Growth in the Next 5 Years
sorted_pe <- unlist(sort(clean_data$Current.PE, decreasing = TRUE))
indices <- unlist(sort(clean_data$Current.PE, decreasing = TRUE, index.return = TRUE)[[2]])
sorted_growth <- clean_data$Expected.Growth.Next.5.Years[indices]
sorted_pe_growth <- data.frame(pe = sorted_pe, growth = sorted_growth)
dat3 <- subset(sorted_pe_growth, sorted_pe < 300)
plot(dat3$pe, dat3$growth, main = "Current PE Ratio vs. Expected Growth", xlab = "Current PE ratio",
    ylab = "Expected growth (next 5 years)", col = rgb(200, 50, 200, 100, maxColorValue = 255),
    pch = 16)
fit2 <- lm(growth ~ pe, data = dat3)
abline(fit2, col = rgb(200, 50, 200, maxColorValue = 255))
```

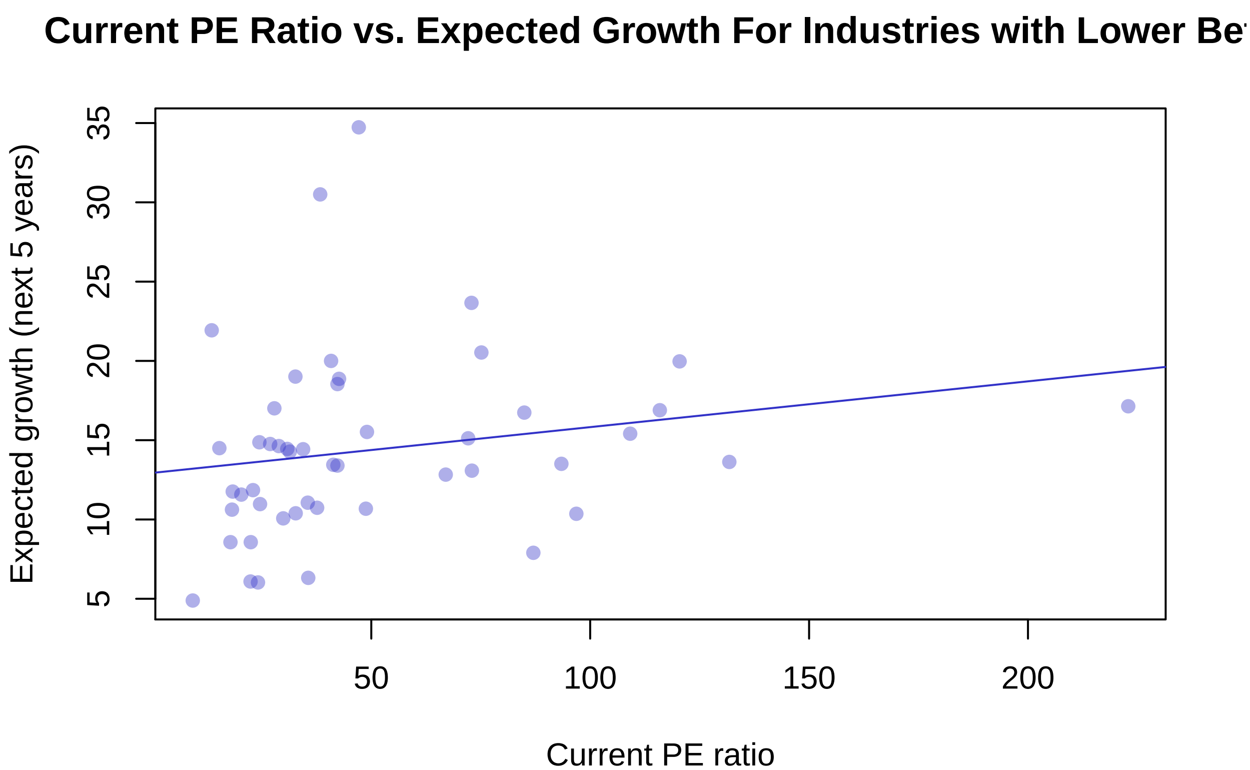



```
summary(fit2)
```

```
##
## Call:
## lm(formula = growth ~ pe, data = dat3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.710  -3.604  -0.699   3.648  22.136
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 15.65566    1.10265  14.198   <2e-16 ***
## pe          -0.01029    0.01397  -0.736    0.464
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.927 on 90 degrees of freedom
## Multiple R-squared:  0.005986,   Adjusted R-squared:  -0.005059
## F-statistic: 0.542 on 1 and 90 DF,  p-value: 0.4635
```

```r
# Account for beta when sorting PE ratios and growth by taking the PE ratios
# and growths of industries with lower beta
sorted_beta <- unlist(sort(clean_data$Average.Unlevered.Beta, index.return = TRUE)[[2]])
beta_pe <- clean_data$Current.PE[sorted_beta]
beta_growth <- clean_data$Expected.Growth.Next.5.Years[sorted_beta]
end <- length(beta_pe)
cutoff <- round(end/2)
beta1 <- data.frame(pe = beta_pe[1:cutoff], growth = beta_growth[1:cutoff])
plot(beta1$pe, beta1$growth, main = paste("Current PE Ratio vs. Expected Growth",
    "For Industries with Lower Beta"), xlab = "Current PE ratio", ylab = "Expected growth (next 5 years
    col = rgb(50, 50, 200, 100, maxColorValue = 255), pch = 16)
fit3 <- lm(growth ~ pe, data = beta1)
abline(fit3, col = rgb(50, 50, 200, maxColorValue = 255))
```




```r
summary(fit3)
```

```
##
## Call:
## lm(formula = growth ~ pe, data = beta1)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.3137 -3.1553 -0.7193  1.2575 20.4314
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.93754    1.34085   9.649 1.25e-12 ***
## pe           0.02887    0.02074   1.392    0.171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.69 on 46 degrees of freedom
## Multiple R-squared:  0.04043,    Adjusted R-squared:  0.01957
## F-statistic: 1.938 on 1 and 46 DF,  p-value: 0.1706
```

```r
# Industries with lower beta
print("Industries with lower beta:")
```

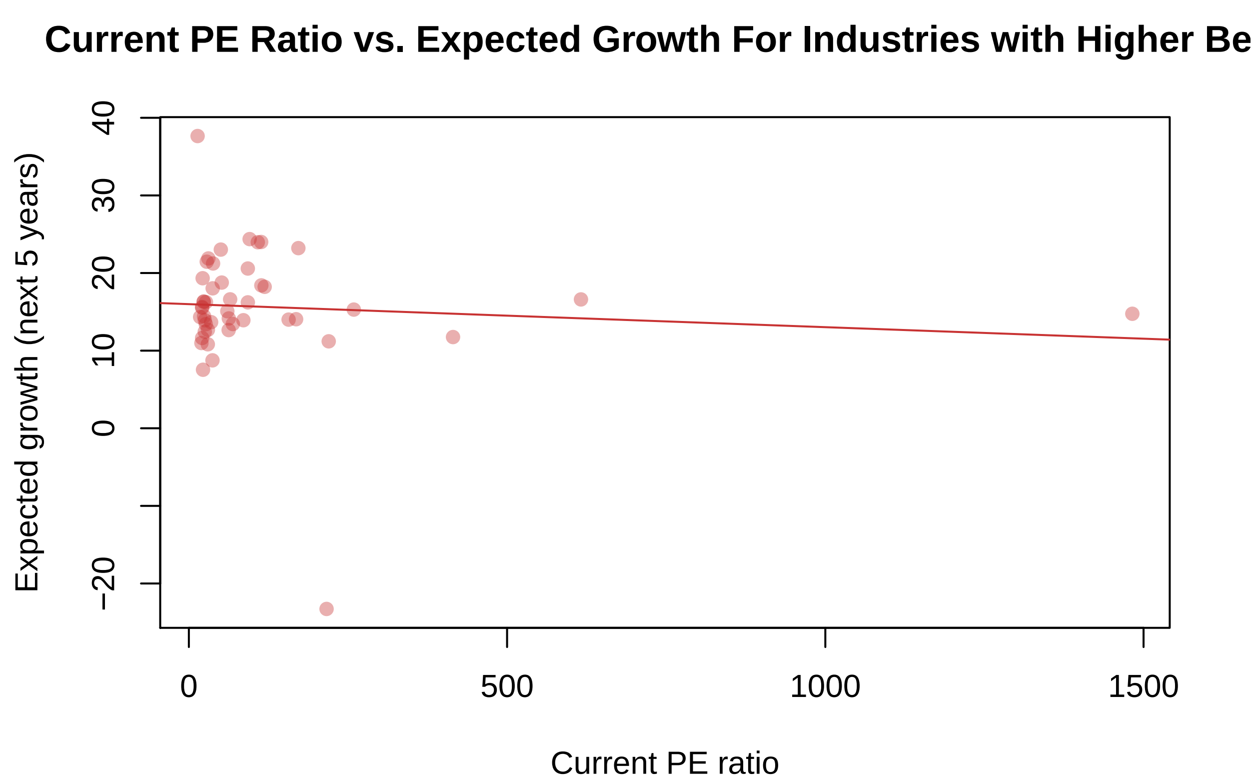
```
## [1] "Industries with lower beta:"
```

```r
print(clean_data$Industry[sorted_beta][1:cutoff])
```

```
##  [1] "Financial Svcs. (Non-bank & Insurance)"
##  [2] "Bank (Money Center)"
##  [3] "Banks (Regional)"
##  [4] "Brokerage & Investment Banking"
##  [5] "Utility (General)"
##  [6] "R.E.I.T."
##  [7] "Telecom (Wireless)"
##  [8] "Power"
##  [9] "Farming/Agriculture"
## [10] "Auto & Truck"
## [11] "Hospitals/Healthcare Facilities"
## [12] "Paper/Forest Products"
## [13] "Air Transport"
## [14] "Rubber& Tires"
## [15] "Oil/Gas Distribution"
## [16] "Green & Renewable Energy"
## [17] "Insurance (Prop/Cas.)"
## [18] "Telecom. Services"
## [19] "Cable TV"
## [20] "Diversified"
## [21] "Packaging & Container"
## [22] "Total Market"
## [23] "Investments & Asset Management"
## [24] "Restaurant/Dining"
## [25] "Chemical (Basic)"
## [26] "Insurance (Life)"
## [27] "Retail (Grocery and Food)"
## [28] "Oil/Gas (Integrated)"
## [29] "Transportation"
## [30] "Utility (Water)"
```

```
## [31] "Insurance (General)"
## [32] "Retail (Distributors)"
## [33] "Food Processing"
## [34] "Real Estate (Development)"
## [35] "Shoe"
## [36] "Advertising"
## [37] "Broadcasting"
## [38] "Coal & Related Energy"
## [39] "Hotel/Gaming"
## [40] "Heathcare Information and Technology"
## [41] "Retail (Automotive)"
## [42] "Retail (General)"
## [43] "Retail (Special Lines)"
## [44] "Apparel"
## [45] "Publshing & Newspapers"
## [46] "Beverage (Alcoholic)"
## [47] "Real Estate (Operations & Services)"
## [48] "Healthcare Products"
```

```
# Account for beta when sorting PE ratios and growth by taking the PE ratios
# and growths of industries with higher beta
beta2 <- data.frame(pe = beta_pe[(cutoff + 1):end], growth = beta_growth[(cutoff +
    1):end])
plot(beta2$pe, beta2$growth, main = paste("Current PE Ratio vs. Expected Growth",
    "For Industries with Higher Beta"), xlab = "Current PE ratio", ylab = "Expected growth (next 5 year
    col = rgb(200, 50, 50, 100, maxColorValue = 255), pch = 16)
fit4 <- lm(growth ~ pe, data = beta2)
abline(fit4, col = rgb(200, 50, 50, maxColorValue = 255))
```



```r
summary(fit4)
```

```
##
## Call:
## lm(formula = growth ~ pe, data = beta2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -38.630  -2.430  -0.310   3.045  21.699
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15.992075   1.282351  12.471 3.34e-16 ***
## pe          -0.002969   0.004976  -0.597    0.554
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.815 on 45 degrees of freedom
## Multiple R-squared:  0.007849,   Adjusted R-squared:  -0.0142
## F-statistic: 0.356 on 1 and 45 DF,  p-value: 0.5537
```

```r
# Industries with higher beta
print("Industries with higher beta:")
```

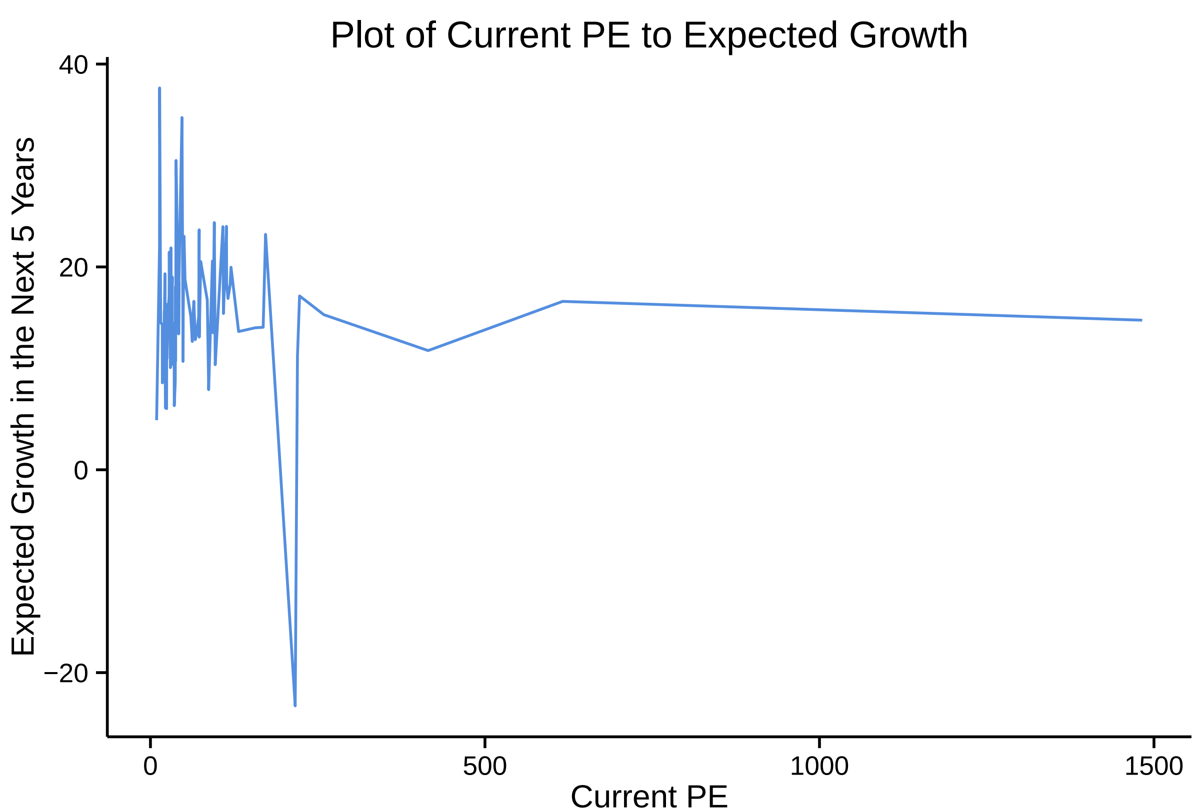
```
## [1] "Industries with higher beta:"
```

```r
print(clean_data$Industry[sorted_beta][(cutoff + 1):end])
```

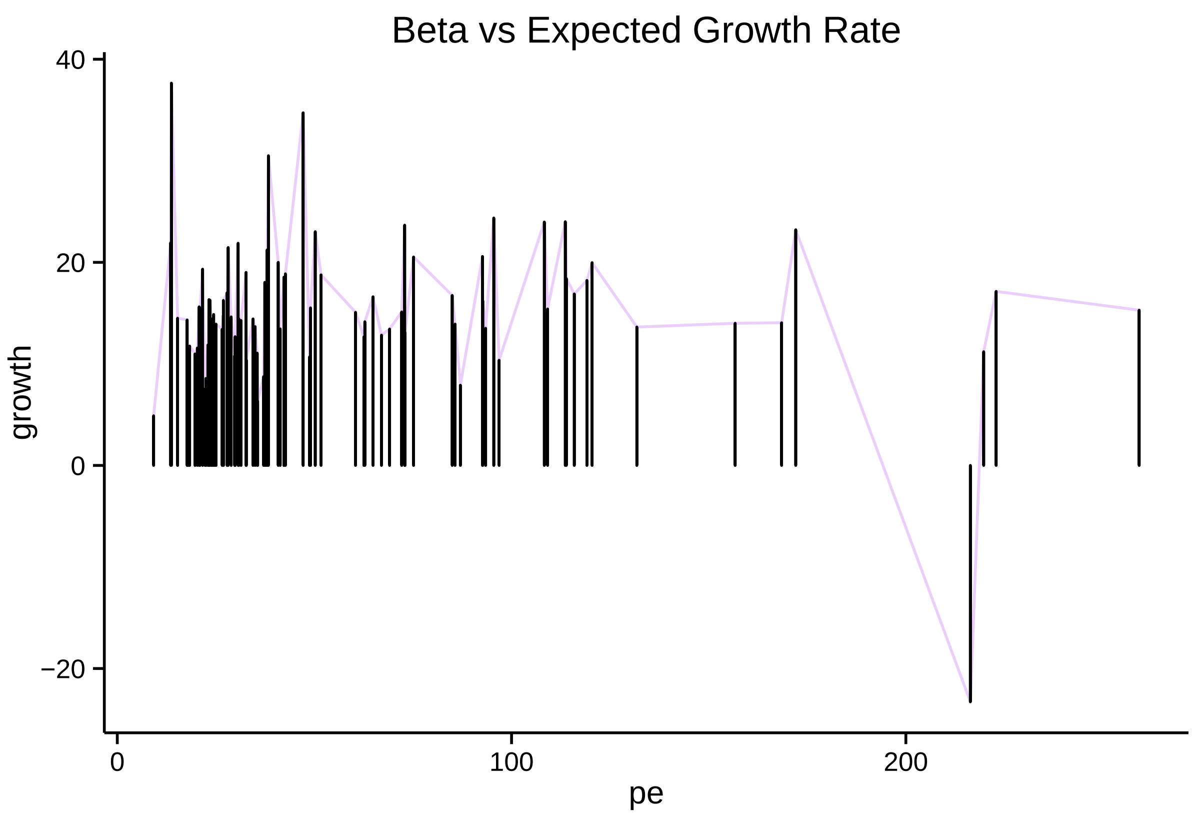
```
##  [1] "Steel"
##  [2] "Chemical (Specialty)"
##  [3] "Healthcare Support Services"
##  [4] "Household Products"
##  [5] "Metals & Mining"
##  [6] "Oil/Gas (Production and Exploration)"
##  [7] "Furn/Home Furnishings"
##  [8] "Homebuilding"
##  [9] "Transportation (Railroads)"
## [10] "Trucking"
## [11] "Building Materials"
## [12] "Environmental & Waste Services"
## [13] "Shipbuilding & Marine"
## [14] "Tobacco"
## [15] "Drugs (Pharmaceutical)"
## [16] "Education"
## [17] "Beverage (Soft)"
## [18] "Chemical (Diversified)"
## [19] "Computer Services"
## [20] "Entertainment"
## [21] "Recreation"
## [22] "Business & Consumer Services"
## [23] "Office Equipment & Services"
## [24] "Electronics (General)"
```

```
## [25] "Information Services"
## [26] "Precious Metals"
## [27] "Aerospace/Defense"
## [28] "Drugs (Biotechnology)"
## [29] "Software (System & Application)"
## [30] "Machinery"
## [31] "Reinsurance"
## [32] "Software (Entertainment)"
## [33] "Auto Parts"
## [34] "Electrical Equipment"
## [35] "Computers/Peripherals"
## [36] "Semiconductor"
## [37] "Semiconductor Equip"
## [38] "Engineering/Construction"
## [39] "Telecom. Equipment"
## [40] "Construction Supplies"
## [41] "Food Wholesalers"
## [42] "Retail (Building Supply)"
## [43] "Software (Internet)"
## [44] "Oilfield Svcs/Equip."
## [45] "Electronics (Consumer & Office)"
## [46] "Retail (Online)"
## [47] "Real Estate (General/Diversified)"
```

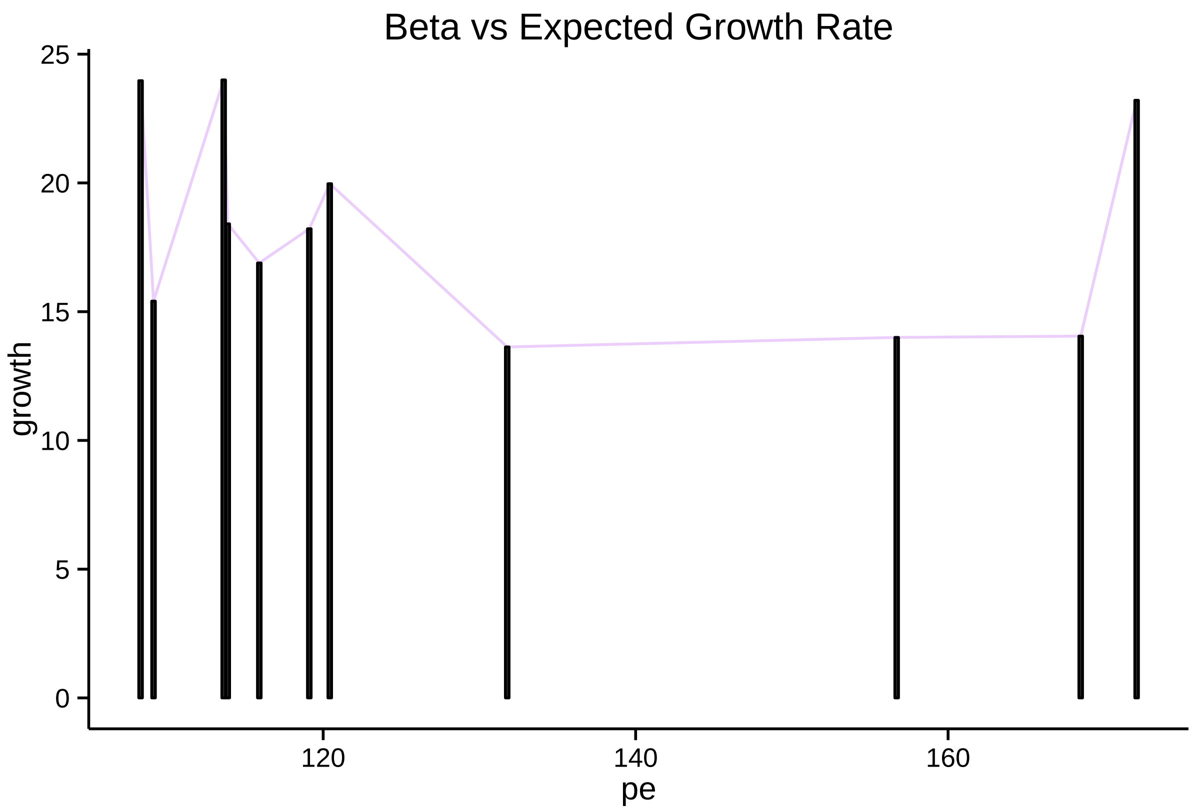
```r
# Graph Current PE to Expected Growth in the Next 5 Years
ggplot(data = clean_data, aes(x = Current.PE, y = Expected.Growth.Next.5.Years)) +
    geom_line(color = "#548edf") + xlab("Current PE") + ylab("Expected Growth in the Next 5 Years") +
    ggtitle("Plot of Current PE to Expected Growth") + theme_classic()
```



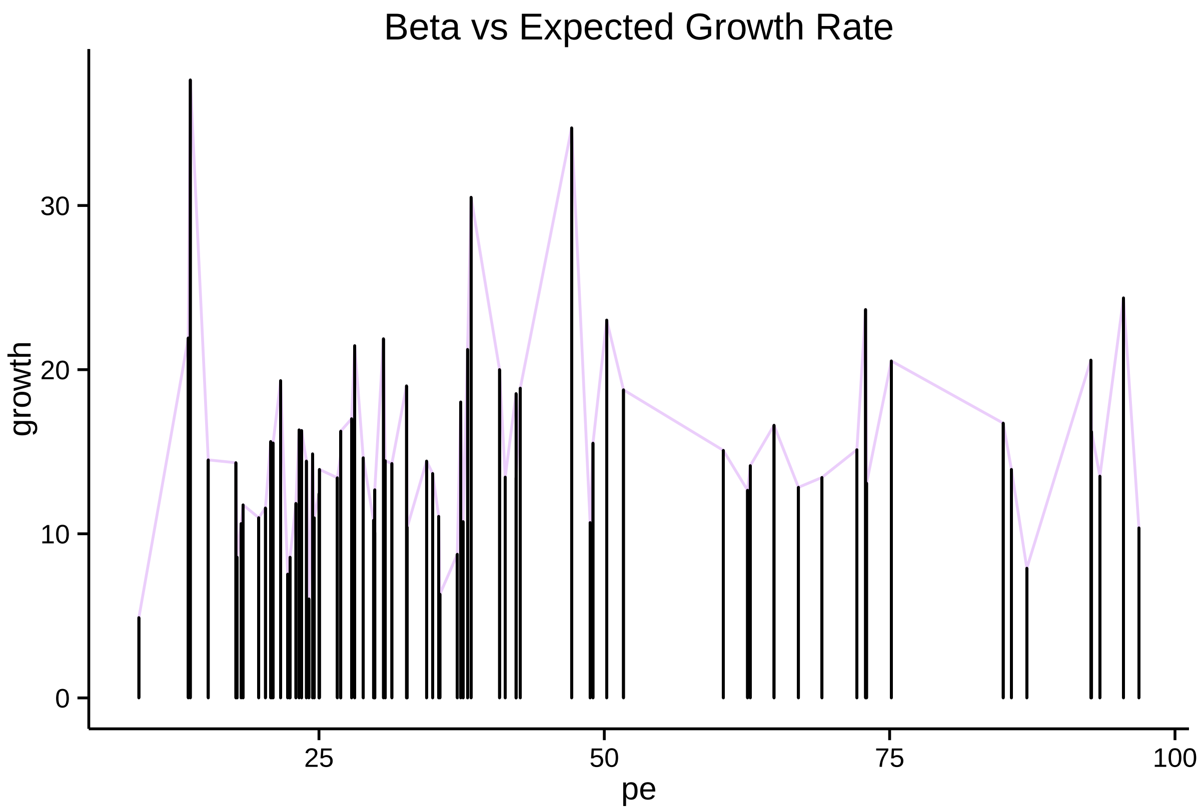
```
# Make a histogram of Current PE to Expected Growth for industries with
# growth, g, fulfilling g < 300
ggplot(dat3, aes(x = pe, y = growth)) + geom_line(color = "#ebcefb") + geom_bar(position = "identity",
    stat = "identity", color = "#000000") + ggtitle("Beta vs Expected Growth Rate") +
    theme_classic()
```



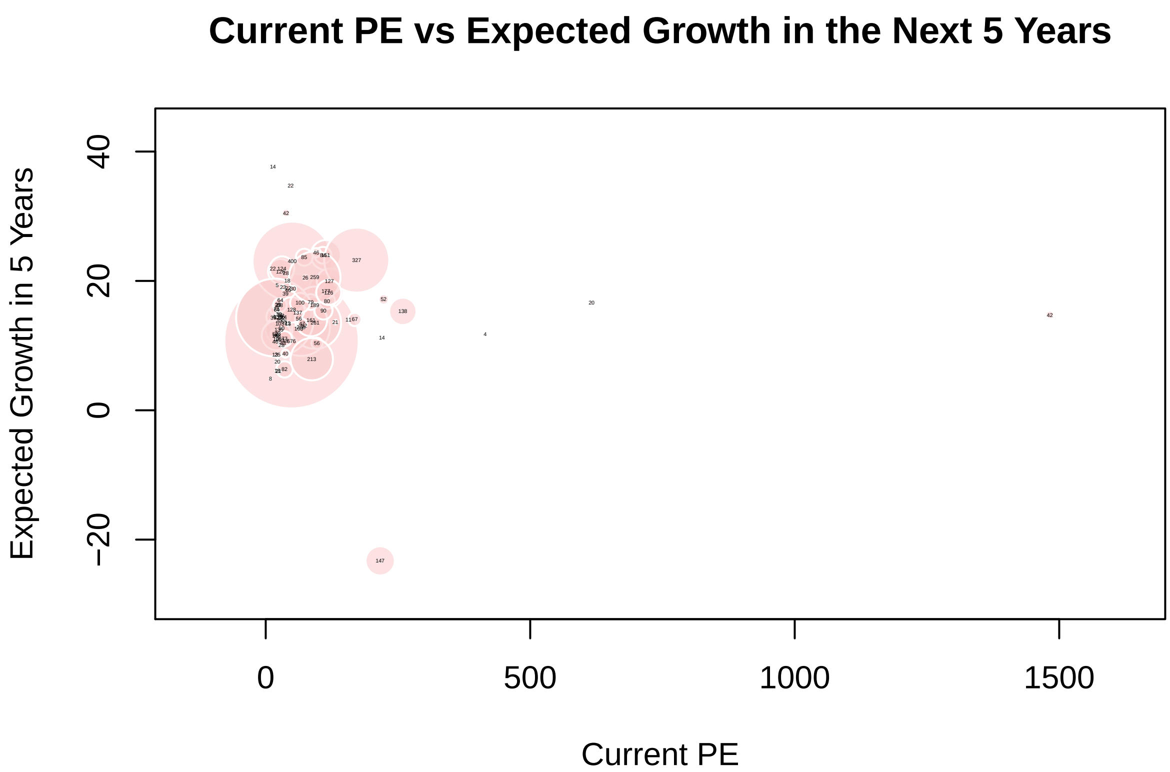

```
# Make a histogram of Current PE to Expected Growth for industries with
# growth, g, fulfilling 100 <= g < 200
dat4 <- subset(sorted_pe_growth, sorted_pe >= 100 & sorted_pe < 200)
ggplot(dat4, aes(x = pe, y = growth)) + geom_line(color = "#ebcefb") + geom_bar(position = "identity",
    stat = "identity", color = "#000000") + ggtitle("Beta vs Expected Growth Rate") +
    theme_classic()
```




```
# Make a histogram of Current PE to Expected Growth for industries with
# growth, g, fulfilling g < 100
dat5 <- subset(sorted_pe_growth, sorted_pe < 100)
ggplot(dat5, aes(x = pe, y = growth)) + geom_line(color = "#ebcefb") + geom_bar(position = "identity",
    stat = "identity", color = "#000000") + ggtitle("Beta vs Expected Growth Rate") +
    theme_classic()
```

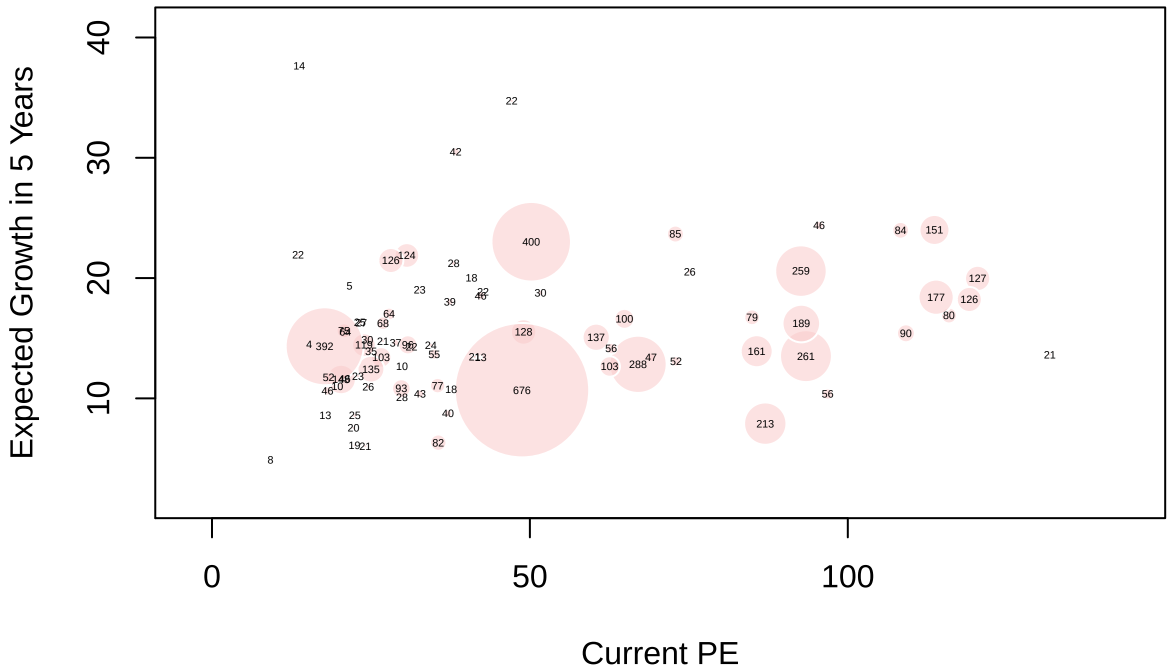



```r
# Make bubble plots of Current PE and Expected Growth in the Next 5 Years
# for different industries
symbols(industries_only$Current.PE, industries_only$Expected.Growth.Next.5.Years,
    circles = industries_only$Number.of.Firms, inches = 0.35, fg = "white",
    bg = "#faccccc8d", xlab = "Current PE", ylab = "Expected Growth in 5 Years",
    main = "Current PE vs Expected Growth in the Next 5 Years")
text(industries_only$Current.PE, industries_only$Expected.Growth.Next.5.Years,
    labels = industries_only$Number.of.Firms, cex = 0.2)
```




```r
# Make bubble plots of Current PE and Expected Growth in the Next 5 Years
# for different industries accounting for outliers
industry_pe <- unlist(sort(industries_only$Current.PE, decreasing = TRUE))
industry_indices <- unlist(sort(industries_only$Current.PE, decreasing = TRUE,
    index.return = TRUE)[[2]])
industry_growth <- industries_only$Expected.Growth.Next.5.Years[industry_indices]
industry_num_firms <- industries_only$Number.of.Firms[industry_indices]
industry_pe_growth <- data.frame(pe = industry_pe, growth = industry_growth,
    num_firms = industry_num_firms)
dat6 <- subset(industry_pe_growth, industry_pe < 150)
symbols(dat6$pe, dat6$growth, circles = dat6$num_firms, inches = 0.35, fg = "white",
    bg = "#faccc8d", xlab = "Current PE", ylab = "Expected Growth in 5 Years",
    main = paste("Current PE Less Than 150 vs", "Expected Growth in the Next 5 Years"))
text(dat6$pe, dat6$growth, labels = dat6$num_firms, cex = 0.3)
```

Current PE Less Than 150 vs Expected Growth in the Next 5 Years



————- *Anais* ——————- *Analysis of relationship between Average Unlevered Beta and Expected Growth for the next 5 years:*

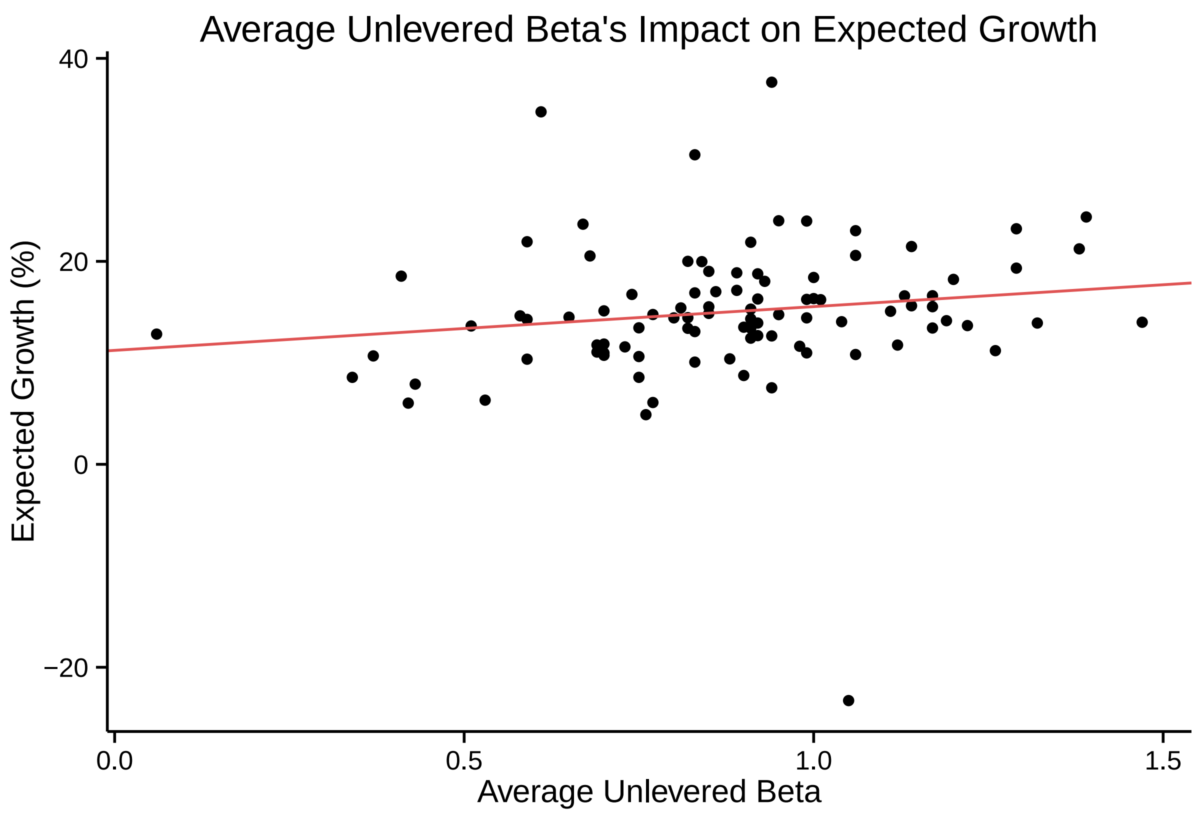
To investigrate the relationship between beta and expected growth over the next 5 years, the data required further cleaning. In order to run a regression between the two variables, the "%" signs in the values of the 5 year expected growth column were removed and the values were converted to numeric values. Using these values, a clean_data column "exp_growth" was aded to the dataframe "clean_data".

```
# Graph the scatterplot of Unlevered Beta's impact on Expected Growth when
# accounting for their regression
reg1 <- lm(Expected.Growth.Next.5.Years ~ Average.Unlevered.Beta, data = clean_data)
summary(reg1)
```

```
##
## Call:
## lm(formula = Expected.Growth.Next.5.Years ~ Average.Unlevered.Beta,
##     data = clean_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.034  -2.919  -0.313   2.432  22.370
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)              11.229      2.579   4.355 3.42e-05 ***
## Average.Unlevered.Beta    4.310      2.823   1.527     0.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.761 on 93 degrees of freedom
```

```
## Multiple R-squared:  0.02445,    Adjusted R-squared:  0.01396
## F-statistic: 2.331 on 1 and 93 DF,  p-value: 0.1302
```

```
ggplot(clean_data, aes(x = Average.Unlevered.Beta, y = Expected.Growth.Next.5.Years)) +
    geom_point() + xlab("Average Unlevered Beta") + ylab("Expected Growth (%)") +
    geom_abline(aes(slope = unname(coef(reg1)["Average.Unlevered.Beta"]), intercept = unname(coef(reg1)
        color = rgb(223, 84, 84, maxColorValue = 255)) + ggtitle("Average Unlevered Beta's Impact on Ex
    theme_classic()
```



The regression returns the estimate for the coefficient of unlevered beta as 2.804, while the coefficient of the interecpt is 13.847. However, the summary of the regression, shows that regressing unlevered beta on exp_growth does not provide a good linear relationship between the two variables as the coefficient does not have linear statistical significance.

The scatterplot of unlevered beta and expected growth further demonstrates the lack of a linear relationship between the two variables. We can see visually that the correlation between the two variables isn't strong.

Further relationships were investigated as other variables that might be causing a deviation in the coefficient for the regression of unlevered beta and expected growth needed to be controlled for.

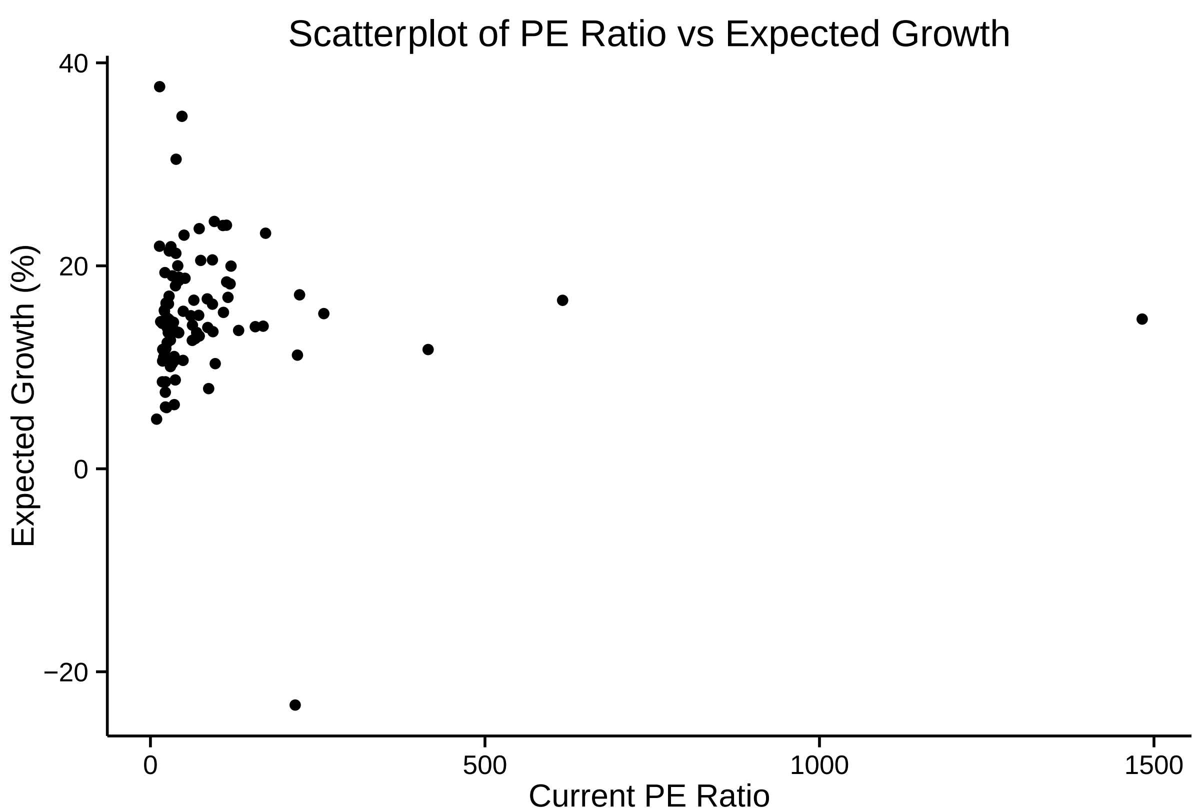
One of these variables is the Price Earnings (P/E) ratio:

```
# Find the multivariate regression of expected growth to unlevered beta and
# current PE ratio
reg2 <- lm(Expected.Growth.Next.5.Years ~ Average.Unlevered.Beta + Current.PE,
    data = clean_data)
summary(reg2)
```

```
##
```

```
## Call:
## lm(formula = Expected.Growth.Next.5.Years ~ Average.Unlevered.Beta +
##     Current.PE, data = clean_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -38.776  -3.056  -0.340   2.482  22.196
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            11.205899   2.588909   4.328 3.81e-05 ***
## Average.Unlevered.Beta  4.552639   2.869431   1.587    0.116
## Current.PE             -0.002266   0.004224  -0.536    0.593
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.788 on 92 degrees of freedom
## Multiple R-squared:  0.0275, Adjusted R-squared:  0.006354
## F-statistic: 1.301 on 2 and 92 DF,  p-value: 0.2773
```

```r
# Graph the scatterplot of unlevered beta's impact on expected growth
ggplot(clean_data, aes(x = Current.PE, y = Expected.Growth.Next.5.Years)) +
    geom_point() + xlab("Current PE Ratio") + ylab("Expected Growth (%)") +
    ggtitle("Scatterplot of PE Ratio vs Expected Growth") + theme_classic()
```



The summary of this regression also demonstrates that there isn't a significant linear relationship between these variables as the only statistically significant coefficient is that of the intercept.

Again, having a look at the relationship visually through the ggplot, we see that there isn't a strong linear relationship and there are outliers present. (#####ASK PETER######)

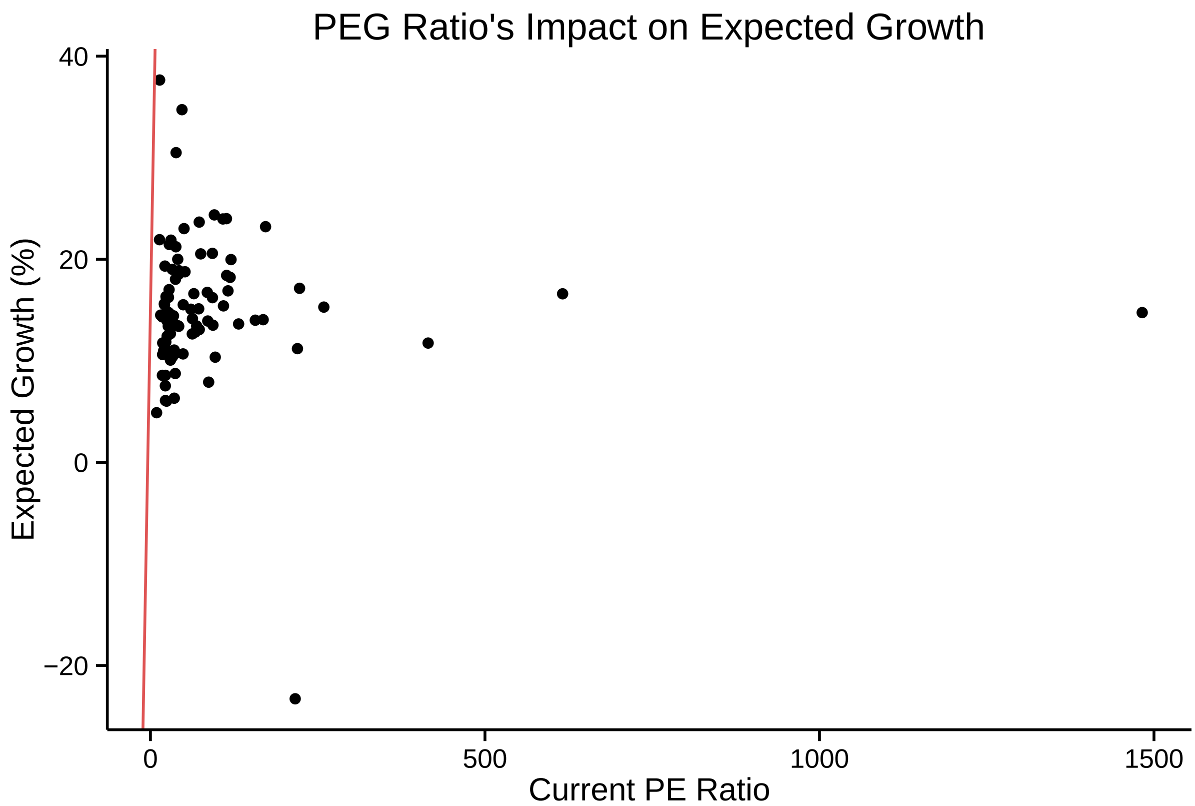
These were interesting findings, which led to the investigration of other related variables, such as the PEG Ratio. The Price/Earnings to Growth (PEG) is a stock's price-to-earnings ratio divided by the growth rate of its earnings for a specified time period. The PEG ratio is used to determine a stock's value while taking the company's earnings growth into account, and is considered to provide a more complete picture than the P/E ratio.

This next regression checks for a linear relationship between PEG ratio and exp_growth, meanwhile retaining PE ratio and unlevered beta in the regression to eliminate any possible effects of confounding factors.

```
# Graph the scatterplot of PEG ratio's impact on expected growth after
# finding the multivariate regression of expected growth in the next 5 years
# to average unlevered beta, the current PE ratio, and the PEG Ratio.  The
# graph only concerns the regression of expected growth to beta.
reg3 <- lm(Expected.Growth.Next.5.Years ~ Average.Unlevered.Beta + Current.PE +
    PEG.Ratio, data = clean_data)
summary(reg3)
```

```
##
## Call:
## lm(formula = Expected.Growth.Next.5.Years ~ Average.Unlevered.Beta +
##     Current.PE + PEG.Ratio, data = clean_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8367 -3.2950 -0.9627  1.8726 19.3578
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            15.1592144  2.2927365   6.612 2.63e-09 ***
## Average.Unlevered.Beta  3.6949760  2.2136230   1.669 0.098553 .
## Current.PE             -0.0006216  0.0031796  -0.195 0.845458
## PEG.Ratio              -1.9515834  0.5201861  -3.752 0.000311 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.097 on 90 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.1871, Adjusted R-squared:   0.16
## F-statistic: 6.905 on 3 and 90 DF,  p-value: 0.0003087
```

```
ggplot(clean_data, aes(x = Current.PE, y = Expected.Growth.Next.5.Years)) +
    geom_point() + xlab("Current PE Ratio") + ylab("Expected Growth (%)") +
    geom_abline(aes(slope = unname(coef(reg3)["Average.Unlevered.Beta"]), intercept = unname(coef(reg3)
        color = rgb(223, 84, 84, maxColorValue = 255)) + ggtitle("PEG Ratio's Impact on Expected Growth
    theme_classic()
```



The summary of regressing PEG ratio against the expected growth, shows that there is statistical significance to the coefficients. The linear regression returned a coefficient estimate of approximately -1.376 for the PEG ratio and 16.472 for the intercept. The summary shows that the PEG ratio estimate has statistical significance at the ????????? level. (####ASK PETER#####)

Having a look at the relationship through a graphical respresentation, the function abline was used to draw the best fit regression line.

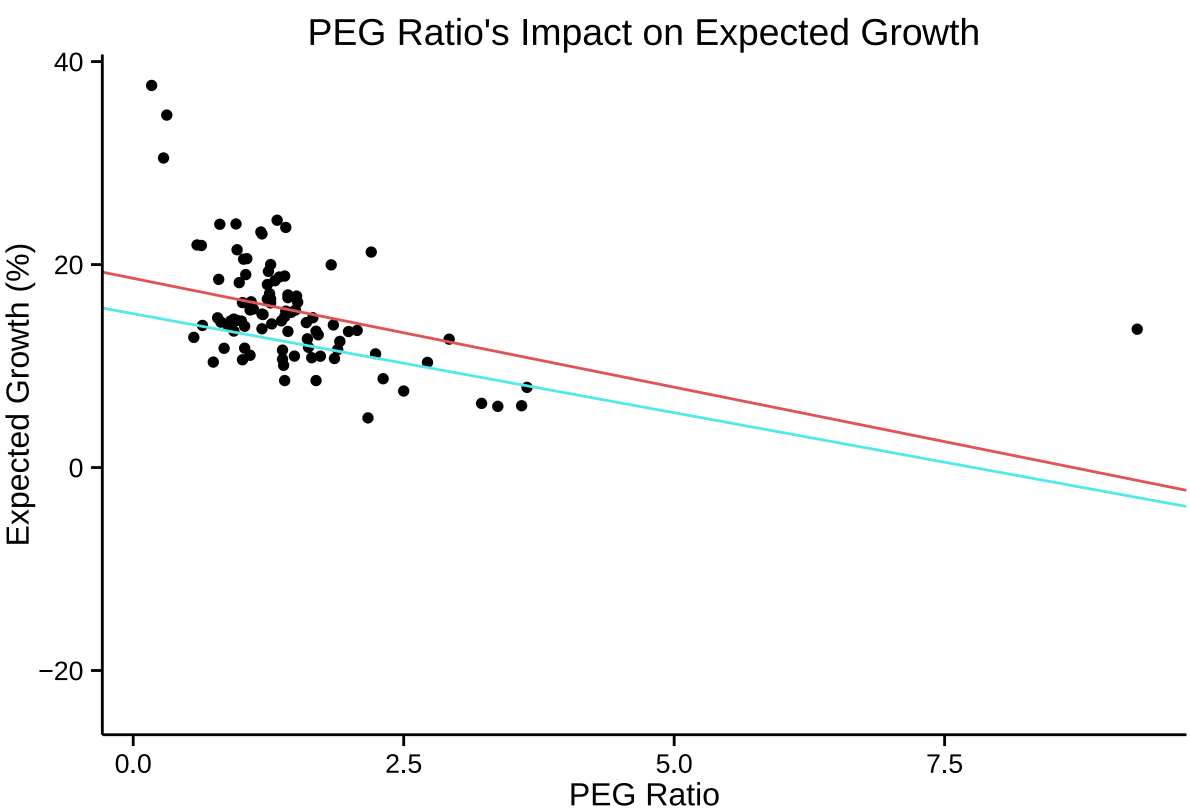
```
# Graph the scatterplot of PEG ratio's impact on expected growth and the
# regression line of expected growth to PEG ratio and the regression line of
# the multivariate regression found in reg3
reg4 <- lm(Expected.Growth.Next.5.Years ~ PEG.Ratio, data = clean_data)
summary(reg4)
```

```
##
## Call:
## lm(formula = Expected.Growth.Next.5.Years ~ PEG.Ratio, data = clean_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.0971 -3.1375 -0.9498  1.9557 19.3720
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  18.6427     0.9274  20.103  < 2e-16 ***
## PEG.Ratio    -2.1455     0.5089  -4.216 5.81e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 5.118 on 92 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.1619, Adjusted R-squared:  0.1528
## F-statistic: 17.77 on 1 and 92 DF,  p-value: 5.811e-05
```

```r
ggplot(clean_data, aes(x = PEG.Ratio, y = Expected.Growth.Next.5.Years)) + geom_point() +
    xlab("PEG Ratio") + ylab("Expected Growth (%)") + geom_abline(aes(slope = unname(coef(reg3)["PEG.Ra
    intercept = unname(coef(reg3)["(Intercept)"])), color = rgb(84, 233, 233,
    maxColorValue = 255)) + geom_abline(aes(slope = unname(coef(reg4)["PEG.Ratio"]),
    intercept = unname(coef(reg4)["(Intercept)"])), color = rgb(223, 84, 84,
    maxColorValue = 255)) + ggtitle("PEG Ratio's Impact on Expected Growth") +
    theme_classic()
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

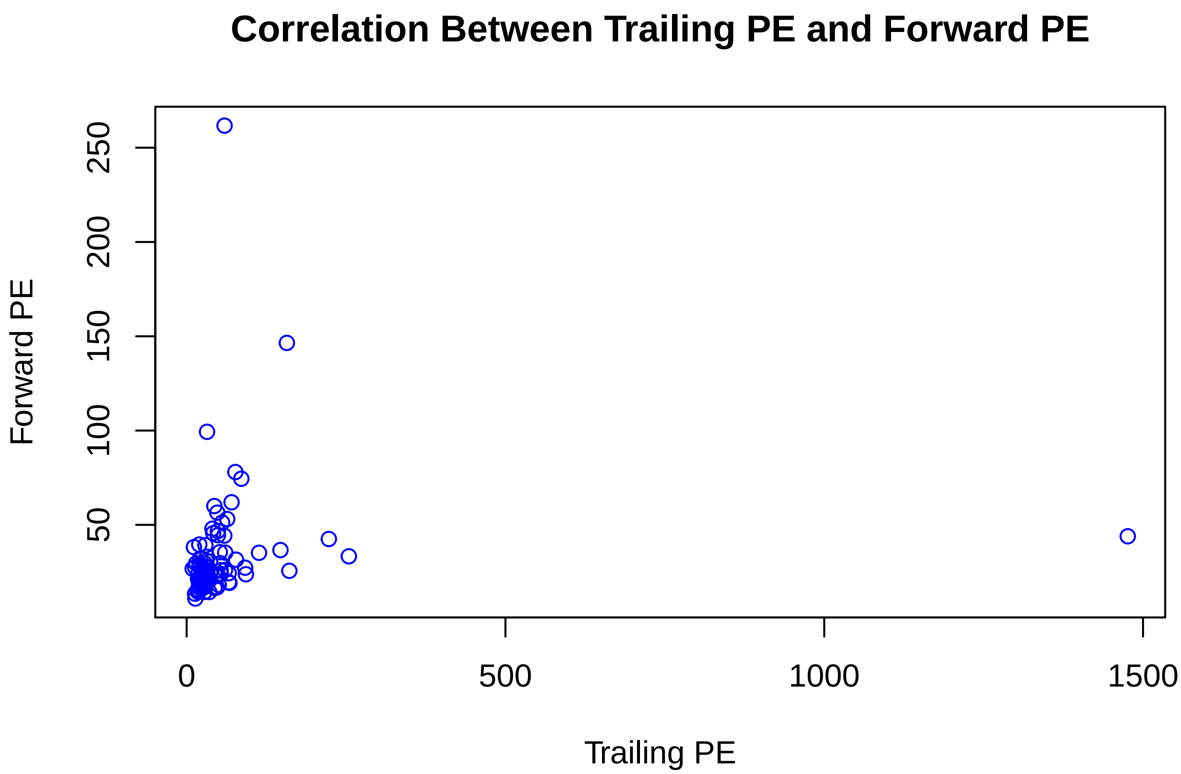


The negative relationship between PEG Ratio and Expected Growth makes sense as the lower the PEG Ratio the higher the growth will be. PEG Ratio is used for stock valuation and is defined as P/E Ratio / Growth. For a good stock you would want a small P/E Ratio and a high denominator as Growth, resulting in a small PEG Ratio.

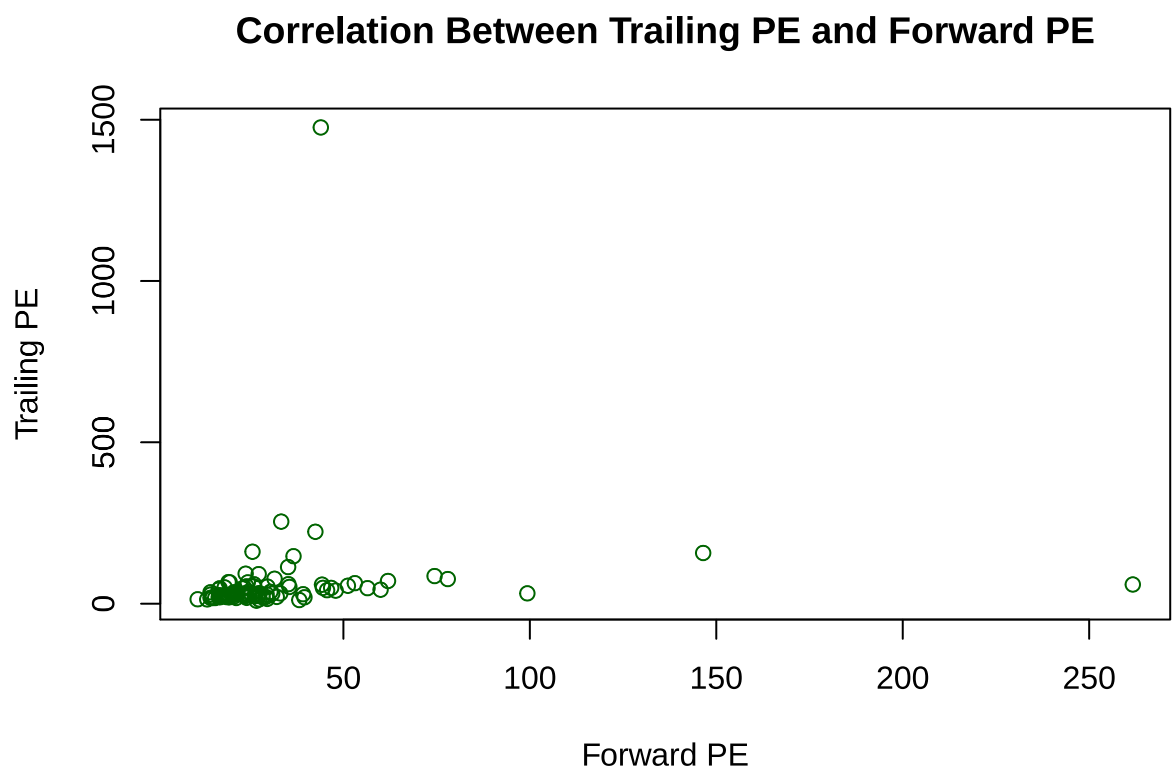
library(scatterplot3d) scatterplot3d(x = clean_dataExpected.growth.in.next.$5.years, clean_{d}ataPEG.Ratio$*, z = clean_data$Average.Unlevered.Beta, angle = 280, scale.y = .3, xlab = "Expected Growth in Next 5 Years", ylab = "PE Ratio", zlab = "Beta")*

Although there is a wide range of industries that have various values of expected growth (thus showing no correlation along with beta and PE ratio), we are still able to see the effects of beta on PE Ratio. As beta increases, PE ratio decreases.

```
plot(clean_data$Trailing.PE, clean_data$Forward.PE, xlab = "Trailing PE",
ylab = "Forward PE", main = "Correlation Between Trailing PE and Forward PE", col = 'blue')
```



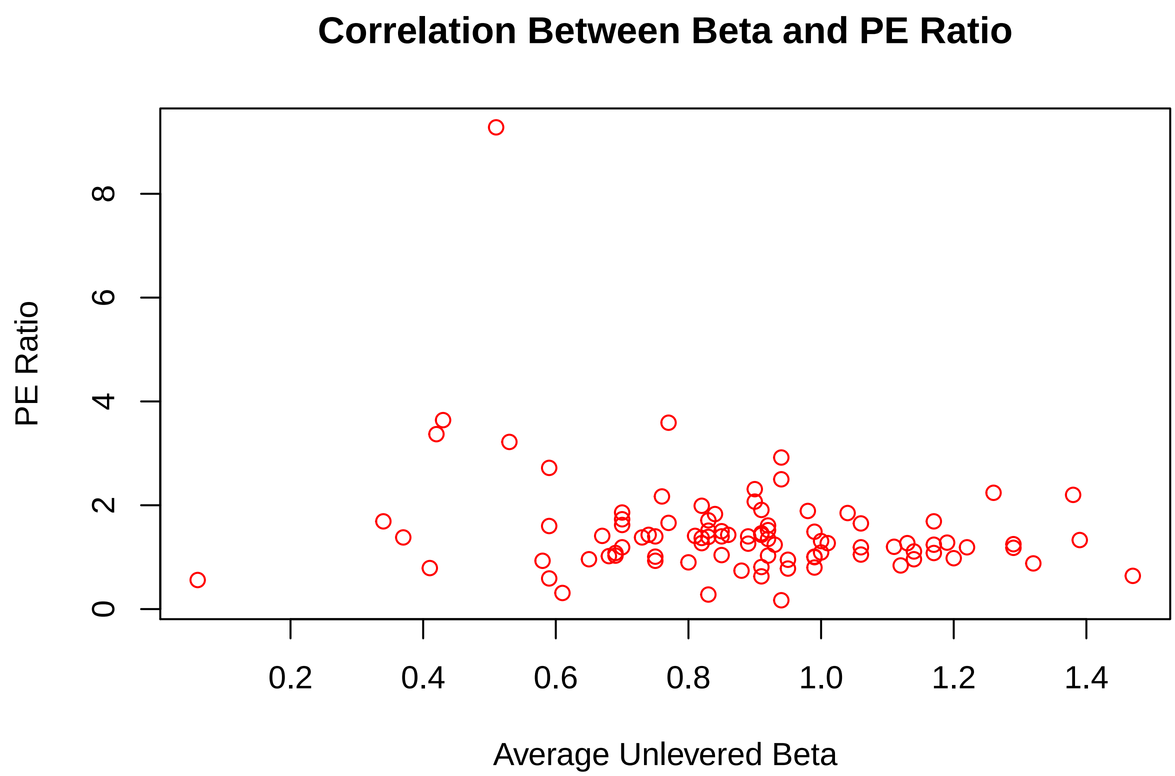

```
plot(clean_data$Forward.PE, clean_data$Trailing.PE, xlab = "Forward PE",
     ylab = "Trailing PE", main = "Correlation Between Trailing PE and Forward PE", col = 'dark green')
```



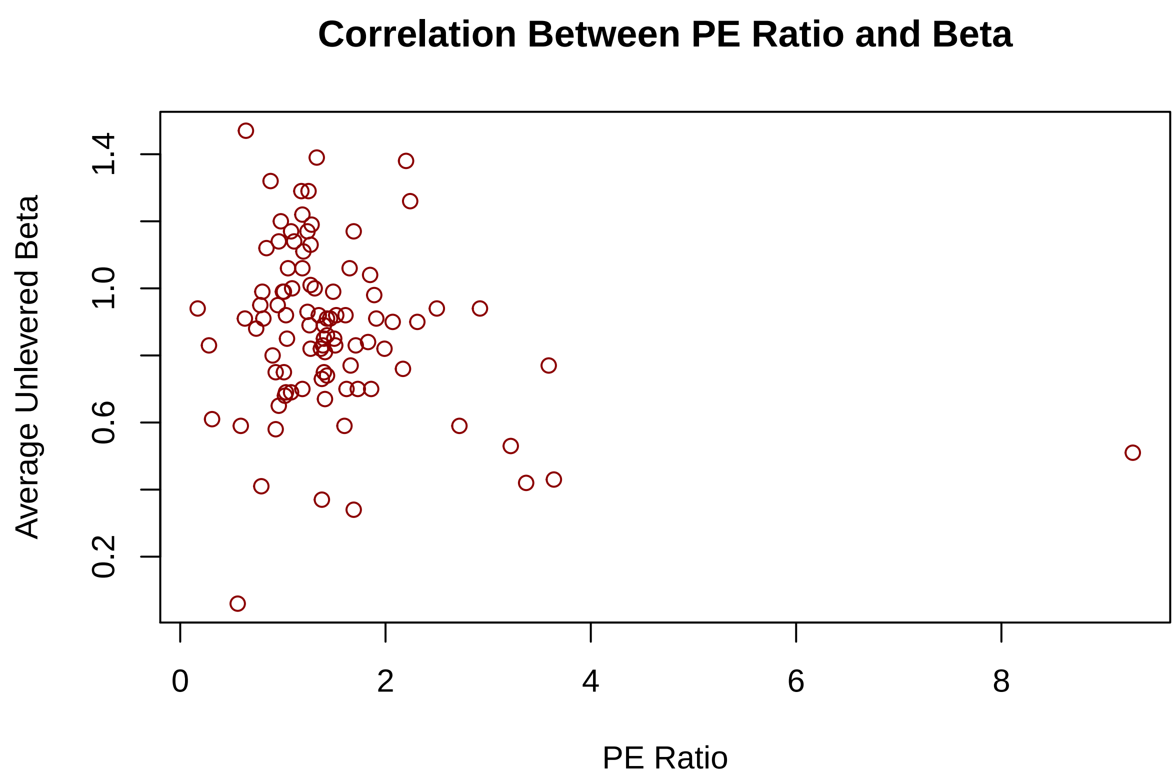
Analysis: Based on the graph that is depicted by the code above, there is a positive correlation between the trailing price earnings and the forward price earnings. This means that the forward price earnings increases as the trailing price earnings increases, and vise verse (though the correlation when the variables are swtiched is not that strong, meaning that the positive impact on the forward price earnings on the trailing price earnings is not as strong as the positive impact that the traling price earnings has on forward price earnings).

Correlation between Average Unlevered Beta and PE ratio

```
plot(clean_data$Average.Unlevered.Beta, clean_data$PEG.Ratio, xlab = "Average Unlevered Beta",
ylab = "PE Ratio", main = "Correlation Between Beta and PE Ratio", col = 'red')
```



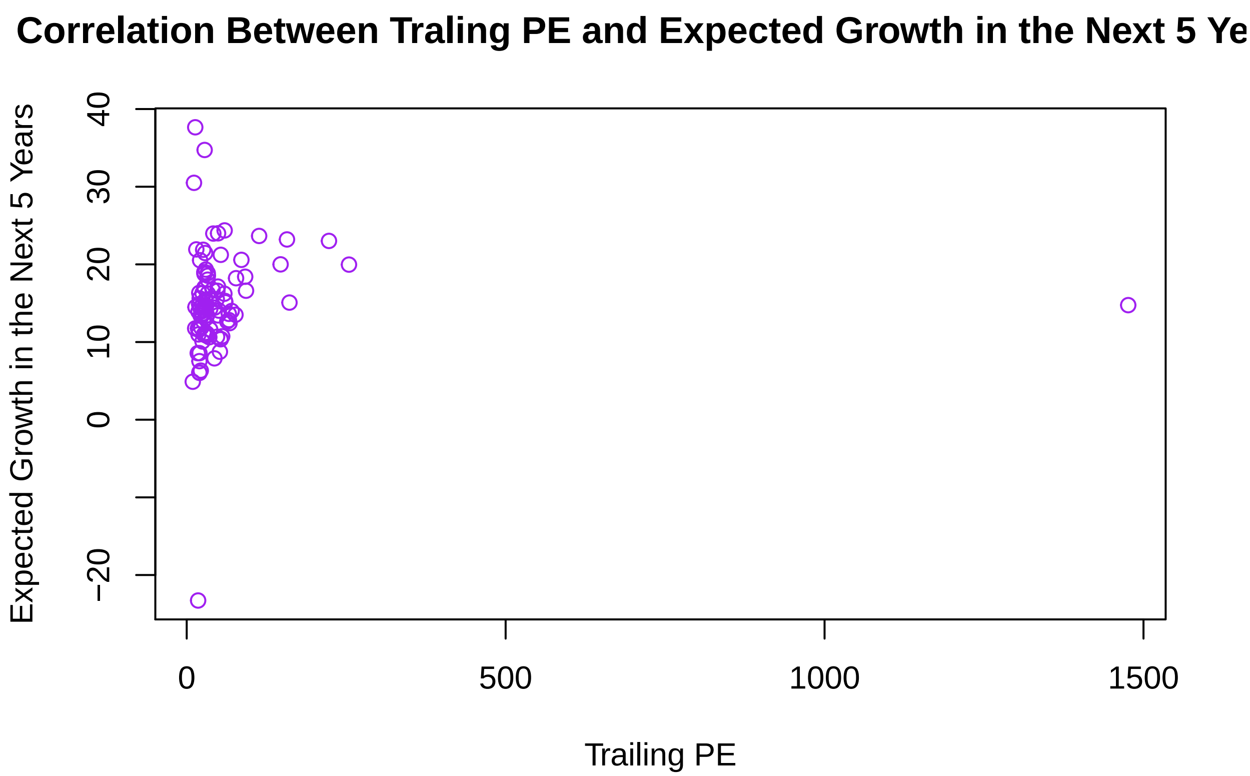

```r
plot(clean_data$PEG.Ratio, clean_data$Average.Unlevered.Beta, xlab = "PE Ratio",
ylab = "Average Unlevered Beta", main = "Correlation Between PE Ratio and Beta", col = 'dark red')
```



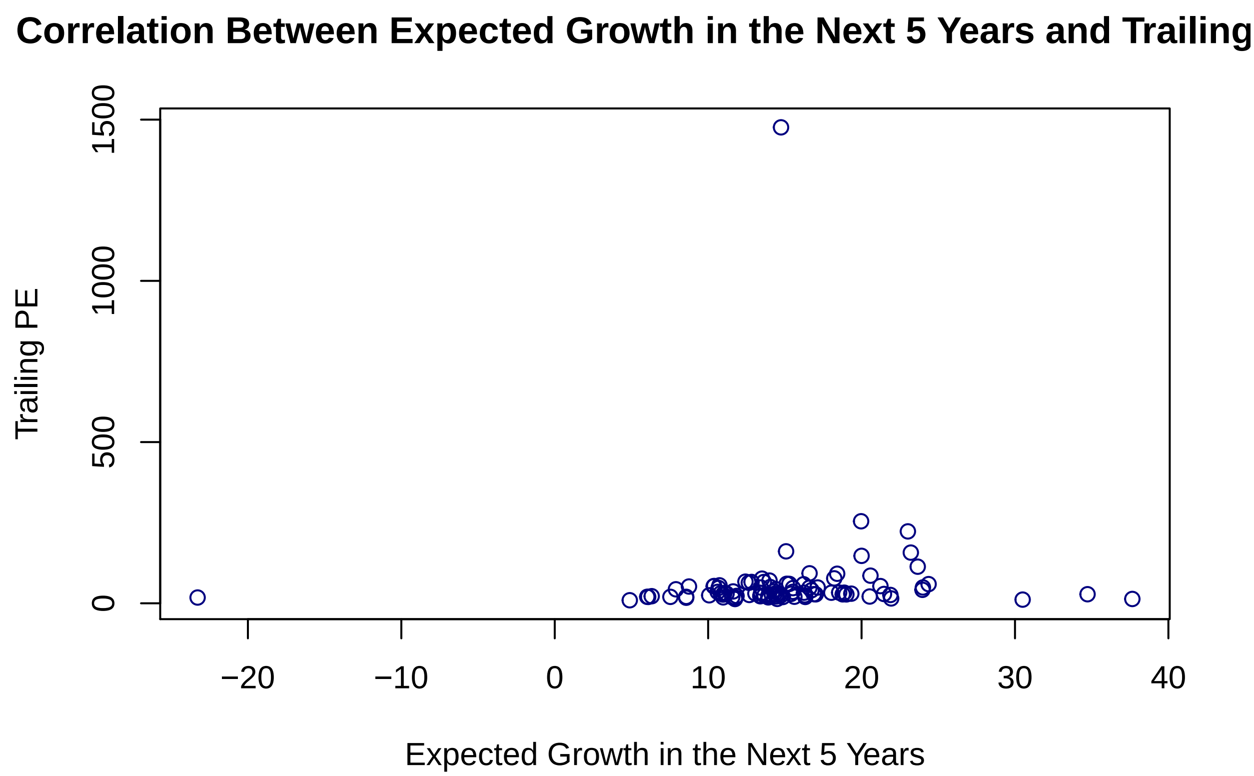
Analysis: Based on the graph that is depicted by the code above, there is almost no correlation between the average unlevered beta and the price earnings (PE) ratio. This means that the average unlevered beta has no *effect on the price earnings ratio in most instances. In contrast, there is a positive correlation between the price earnings ratio and the average unlevered beta. This means that the PE ratio has a positive effect of the average unlevered beta.*

Correlation between trailing PE and Expected growth in the next 5 years.

```
plot(clean_data$Trailing.PE, clean_data$Expected.Growth.Next.5.Years, xlab = "Trailing PE",
ylab = "Expected Growth in the Next 5 Years", main = "Correlation Between Traling PE and Expected Growt
```



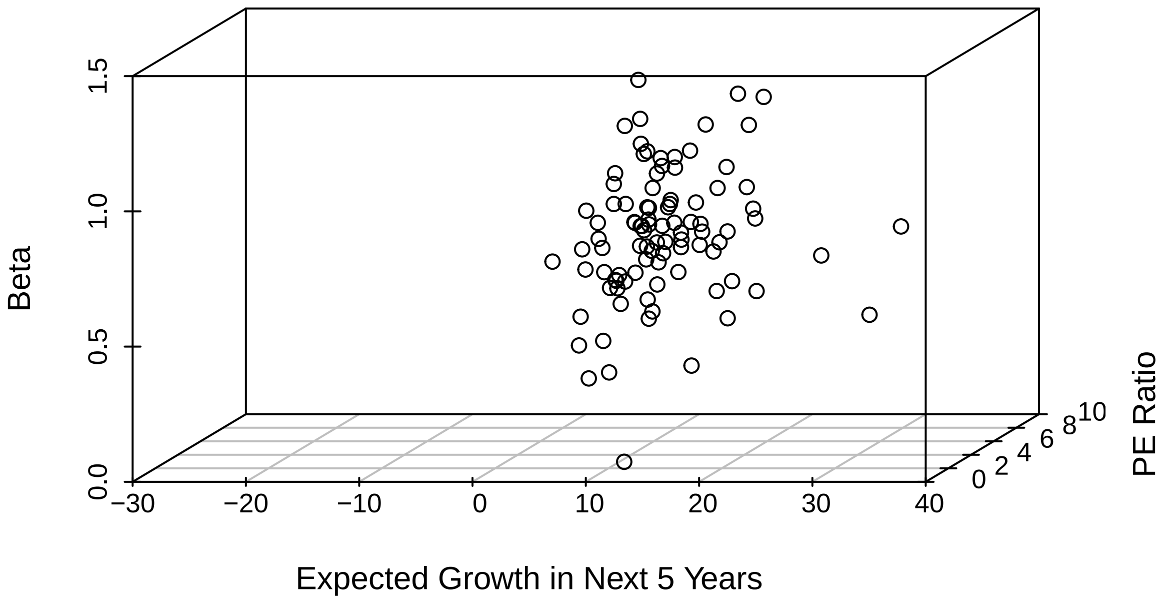

```
plot(clean_data$Expected.Growth.Next.5.Years, clean_data$Trailing.PE, xlab = "Expected Growth in the Ne
ylab = "Trailing PE", main = "Correlation Between Expected Growth in the Next 5 Years and Trailing PE",
```



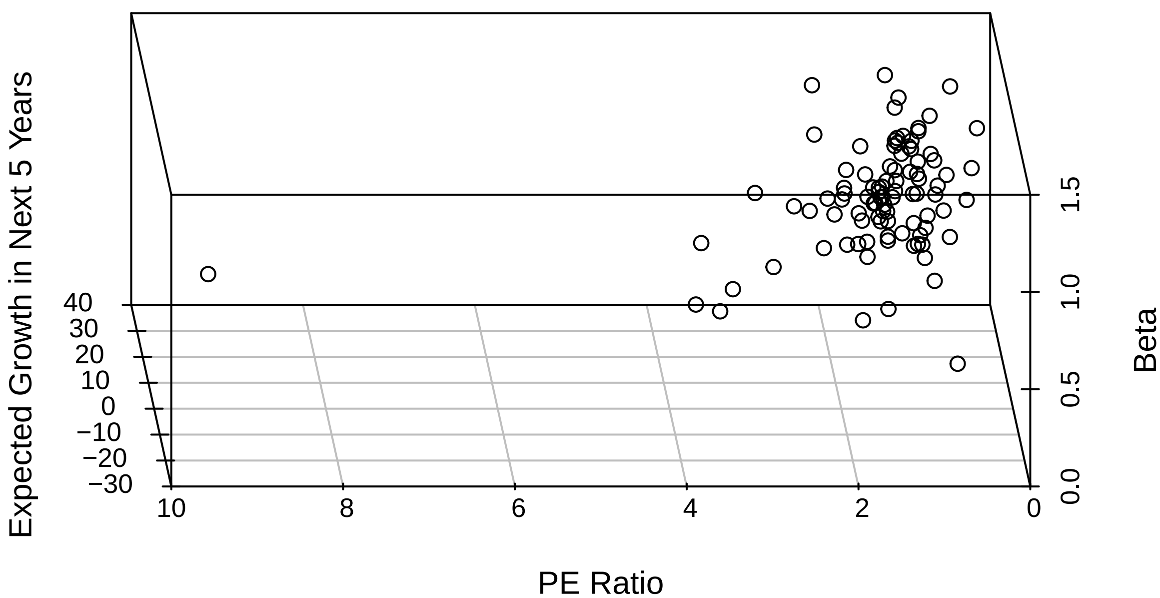
Based on the graphs above, there are no correlations to be found between trailing PE and the expected growth in the next five years. In other words, the trailing PE has no effect on the expected growth in the next five years, and vise versa.

Our Hypothesis: Tech industries show higher growth and higher beta, and therefore lower P/E ratios.

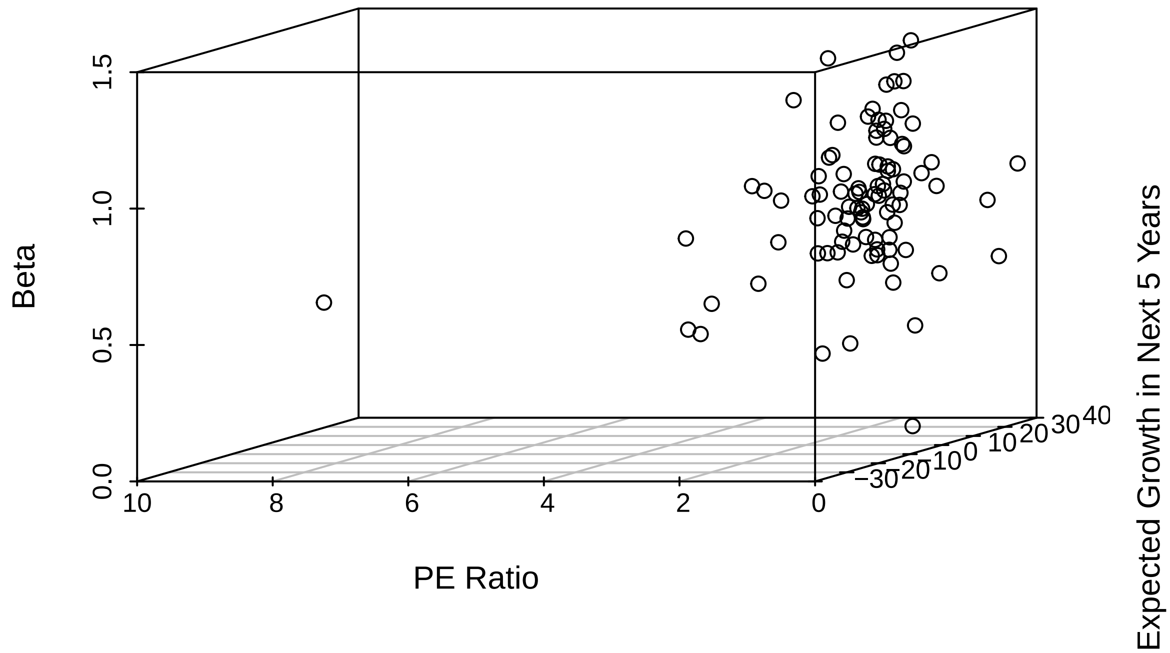
```
scatterplot3d(x = clean_data$Expected.Growth.Next.5.Years,
              clean_data$PEG.Ratio, z = clean_data$Average.Unlevered.Beta,
              angle = 30, scale.y = .3,
              xlab = "Expected Growth in Next 5 Years",
              ylab = "PE Ratio", zlab = "Beta")
```



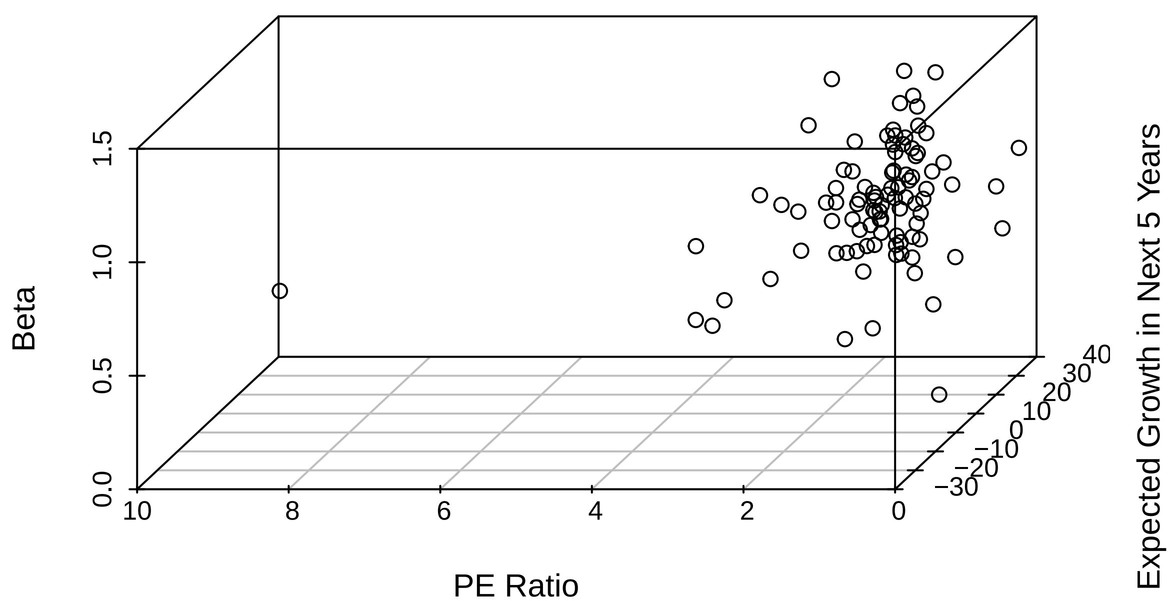

```
# Look at PE Ratio here:
scatterplot3d(x = clean_data$Expected.Growth.Next.5.Years,
              clean_data$PEG.Ratio, z = clean_data$Average.Unlevered.Beta,
              angle = 280, scale.y = .3,
              xlab = "Expected Growth in Next 5 Years",
              ylab = "PE Ratio", zlab = "Beta")
```



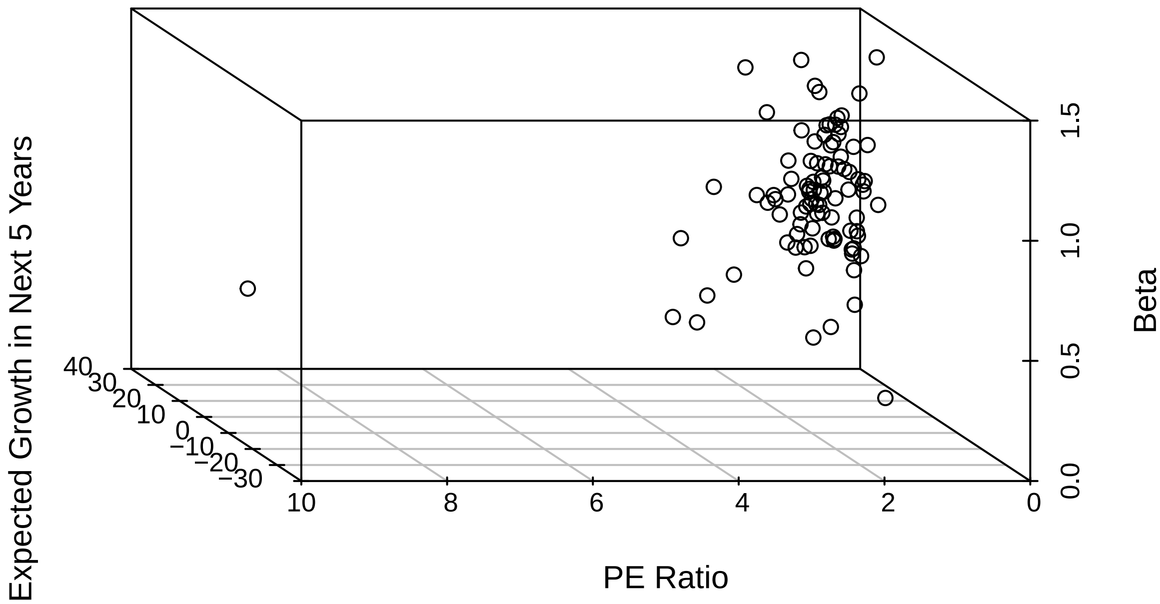

```
# Multiple angles
scatterplot3d(x = clean_data$Expected.Growth.Next.5.Years,
              y = clean_data$PEG.Ratio,
              z = clean_data$Average.Unlevered.Beta,
              angle = 200, scale.y = .3,
              xlab = "Expected Growth in Next 5 Years",
              ylab = "PE Ratio", zlab = "Beta")
```



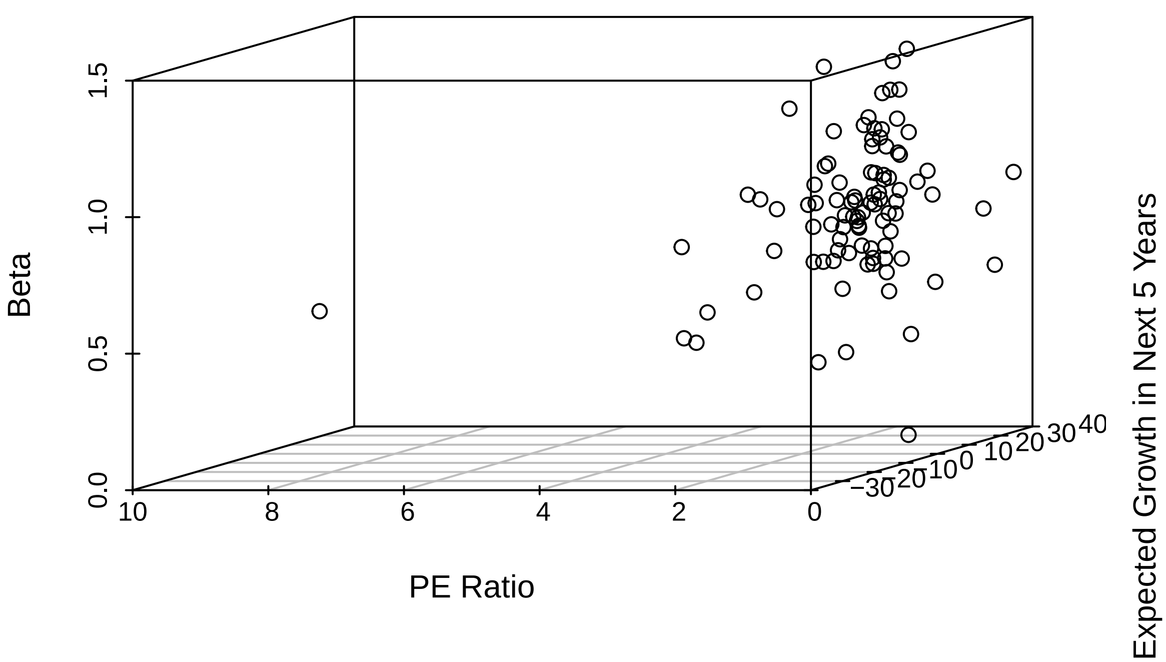

```r
scatterplot3d(x = clean_data$Expected.Growth.Next.5.Years,
              y= clean_data$PEG.Ratio,
              z = clean_data$Average.Unlevered.Beta,
              angle = 230, scale.y = .3,
              xlab = "Expected Growth in Next 5 Years",
              ylab = "PE Ratio", zlab = "Beta")
```




```r
scatterplot3d(x = clean_data$Expected.Growth.Next.5.Years,
              y = clean_data$PEG.Ratio,
              z = clean_data$Average.Unlevered.Beta,
              angle = -400, scale.y = .3,
              xlab = "Expected Growth in Next 5 Years",
              ylab = "PE Ratio", zlab = "Beta")
```




```r
scatterplot3d(x = clean_data$Expected.Growth.Next.5.Years,
              y= clean_data$PEG.Ratio,
              z = clean_data$Average.Unlevered.Beta,
              angle = 200, scale.y = .3,
              xlab = "Expected Growth in Next 5 Years",
              ylab = "PE Ratio", zlab = "Beta")
```



Although there is a wide range of industries that have various values of expected growth (thus showing no correlation along with beta and PE ratio), we are still able to see the effects of beta on PE Ratio. As beta increases, PE ratio decreases.

Findings

Conclusion

write.